UNIVERSITY OF CALIFORNIA,
IRVINE

Ground Motion Simulation Validation through Seismic Performance Assessment of Structural Systems

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Huda Munjy

Dissertation Committee:
Associate Professor Farzin Zareian, Chair
Associate Professor Anne Lemnitzer
Professor Babak Shahbaba

2019

DEDICATION

To

My Family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like express my gratitude to my advisor, Dr. Farzin Zareian, for continuously motivating me, inspiring me and educating me throughout my PhD journey. This dissertation could not have been accomplished without his wisdom, support and guidance. I would like to extend my appreciation to my committee members, Dr. Anne Lemnitzer, who has been a constant source of support and encouragement, and Dr. Babak Shahbaba, whose strong mentorship and keen insight in statistics has helped me throughout my work.

I would like to extend a special thank you to Raslyn Rendon, who has taught me to face my challenges head on and always strive to reach my full potential. I would like to thank my former and current colleagues including Dr. Omid Esmaili, Dr. Bahareh Mobasher, Dr. Peng Zhong, Marta De Bortoli, Dr. Pablo Torres, and Rachelle Habchi. To Jawad Fayaz and Yijun Xiang, I thank you both for always making my visits to the lab full of discovery, ideas, and laughter.

Finally, I would like to thank my family from the bottom of my heart for supporting me throughout my entire educational journey. Zainab and Hibba—thank you for being my sisters, my best friends, and examples of strength, intelligence and confidence to strive to emulate. Luma, you are more than just my big sister—thank you for caring for me, teaching me, and guiding me at every turning point in my life. My parents, my role model civil engineers, without you I could not have dreamed of being where I am today. Mama—thank you for always believing in me, even when I didn't believe in myself. Baba—thank you for teaching me everything I know today and for inspiring me to strive to be just like you. Lastly, I would like to extend my most heartfelt thanks to my husband, Danyal, who has been by my side through it all, who has supported me through every obstacle, and who has shown me how to not only dream the biggest dreams, but to make them reality. Thank you.

CURRICULUM VITAE

Huda Munjy, PhD, EIT

Education

August 2019	**University of California, Irvine** Henry Samueli School of Engineering	**PhD, Civil and Environmental Engineering** Emphasis in Structural Engineering
2011-2013	**California State University, Fresno** Lyles College of Engineering Fresno, California	**Master of Science, Civil and Environmental Engineering** Graduated with distinction GPA: 4.0 Emphasis in Structural Engineering and Geotechnical Engineering
2008-2011	**California State University, Fresno** Lyles College of Engineering Fresno, CA	**Bachelor of Science, Civil and Environmental Engineering** Magna Cum Laude GPA: 3.81 Emphasis in Structural Engineering Water Resources Minor in Mathematics

Certifications

04.2011	**Engineer in Training (EIT)** **National Council of Examiners for Engineering and Surveying (NCEES)**
09.2015	**Mentoring Excellence Program Completion Certification** **University of California, Irvine**

Grants and Scholarships

2019	NSF Travel Grant, $2,250
2013-2019	California State University Chancellor's Doctoral Incentive Program Scholarship, $30,000
2018	Department Diversity Fellowship, University of California, Irvine, $13,000
2018	NSF Travel Grant, $2,500
2012	Building Industry Association Women's Council Scholarship, $5,000
2011	Charles C Buckley Engineering Scholarship, $1,500
2011	James A. Ross Memorial Scholarship, $1,500

Research Experience

09.2013-09.2019 **Graduate Student Researcher**
University of California, Irvine
Advisor: Dr. Farzin Zareian, PhD

- Conducted research in the field of Structural Engineering with an emphasis in Performance Based Earthquake Engineering and Statistics
- Worked with computer programs for civil engineering applications including MATLAB, R Statistical Analysis Software, Python, C++, OpenSees and PACT Loss Estimation Software
- Extensively worked with the UCI Department of Statistics
- Specific research topics include ground motion simulation validation and statistical methods for characterization of engineering demand parameters used in seismic performance assessment of structural systems

01.2011-08.2013 **Research Assistant**
CALTRANS/California State University, Fresno
Advisor: Dr. Fariborz Tehrani, PhD, PE, ENV, SP, PMP, SAP

- CALTRANS funded research project through California State University, Fresno on mechanically stabilized earth retaining walls with light weight aggregate backfill subject to seismic loading
- Aided in grant writing process that included funding for 3 faculty members, 3 graduate student researchers, PLAXIS software for analysis, experimental testing and equipment
- Conducted all analytical modeling and simulations via PLAXIS software
- Authored several detailed technical reports and two journal publications
- Assembled and conducted several research presentations for both CALTRANS staff and California State University, Fresno faculty

Publications

(under review) **Munjy, H.**, Zareian, F., (2019). "Seismic loss estimation based on modeling dependence of peak floor acceleration and maximum interstory drift ratio with Gaussian copulas." *Bulletin of Earthquake Engineering.*

(under review) **Munjy, H**. , Habchi, R., Zareian, F., (2019). "Validation of Simulated Earthquake Ground Motions for Displacement Response of Building and Bridge Structures Based on Intensity and Frequency Content Parameters." *Bulletin of Earthquake Engineering.*

Munjy,H. , Zareian, F., (2018). "Efficient Statistical Approximation of Engineering Demand Parameters in Building Structures." *16th European Conference on Earthquake Engineering.*

Apelian, D., Artis, S., Bennet, J., Burke, P., Cassidy, R., DaSilva, N., Grant, S., Li, G., **Munjy, H.,** Reinkensmeyer, D., Sanders, B., Tran, K., Valdevit, L., Wilkens, J., (2017). "UCI: Excellence- Ideas, Ideals and Impact: The Samueli School Strategic Plan." *University of California, Irvine, Henry Samueli School of Engineering.* http://engineering.uci.edu/files/v28-final-ssoe-str-plan.pdf

Munjy, H., F. Tehrani, M. Xiao, and M. Zoghi., (2014). "A Numerical Simulation on the Dynamic Response of MSE Wall with LWA Backfill." *Numerical Methods in Geotechnical Engineering*: 1147-151.

Xiao, M., Tehrani, F. M., Ledezma, M., Hartman, C., and **Munjy, H.,** (2014). Shake Table Testing and Numerical Analyses of Seismic Responses of Mechanically Stabilized Earth Wall with Tire Derived Aggregate (TDA) Backfill. *Transportation Research Board 93rd Annual Meeting*: No. 14-2825.

Teaching and Mentoring Experience

09.2017-09.2019 **Writing Consultant**
University of California, Irvine
Supervisor: Raslyn Rendon, Director of Alumni Relations and Student Affairs

- Worked as an employee of the Graduate Division
- Assisted graduate students and faculty in editing reports, academic papers, dissertations, grant and scholarship applications and any other academic related writing pieces
- Provided one-on-one meetings where the interactive editing process enhances writing pieces and aids in understanding how to avoid writing related mistakes
- Met with a diverse group of students across multiple disciplines

October 2018 **Graduate Student Orientation Presenter**
University of California, Irvine
Topic: Writing a Graduate Level Research Paper

- Invited by the Graduate Division at UC Irvine to present to incoming graduate students (over 1,000 students)
- Discussed writing essential and resources for students to utilize to improve their writing

October 2018 **Guest Lecturer**
California State University, Fresno: Department of Civil and Environmental Engineering
Course: Research Methods

- Invited by CSU Fresno faculty to design and present a workshop for graduate students regarding how to formulate a research topic and begin writing a research paper
- Provided essential tips for tackling the writing aspect of difficult research projects that aid in honing writing skills and reducing writing-related anxiety

06.2016-present **Instructor**
Advance Academy, Irvine, CA
Supervisor: Jenny Hsieh

- Provide supplementary instruction to high school and college students in a variety of mathematics and science courses including Algebra 1 and 2, Trigonometry, Geometry, Statistics, Calculus, Differential Equations, and Physics
- Work with students in person and via distance learning through online platforms with personally created online learning material
- Instruct junior high to college level students in essay writing workshops geared towards learning report and essay writing skills
- Conduct mentoring sessions and provides guidance for high school students applying to college
- Work with a diverse group of students, including international students and underrepresented minority students where I provide assistance and a personalized learning experience for each student

09.2015-05.2019 **DECADE PLUS Mentor**
University of California, Irvine

Supervisor: Raslyn Rendon, Director of Alumni Relations and Student Affairs

- The Diverse Educations Community and Doctoral Experience (DECADE) program's mission is to increase the number of women and underrepresented minorities receiving doctoral degrees at UCI
- Acted as a mentor for undergraduate students coming from underrepresented minorities and first generation college students
- Worked independently and in a team as a part of a highly structured mentoring program
- Assigned with the main task of following and tracking the progress of six freshman at UCI including required weekly individual and group meetings and monthly full program meetings

03.2014-05.2014 **Soils Engineering Reader**
University of California, Irvine
Course Instructor: Dr. Anne Lemnitzer, PhD

01.2014-03.2014 **Strength of Materials Reader**
University of California, Irvine
Course Instructor: Dr. Farzin Zareian, PhD

08.2012-05.2013 **Teaching Assistant**
California State University, Fresno
Advisor: Dr. Fariborz Tehrani, PhD, PE, ENV, SP, PMP, SAP

- Teaching assistant for Timber Design Course
- Developed and presented lectures series for students as well as assisted in examination and evaluation development
- Assisted professor in grading of homework and assignments for the course
- Held sessions for students to ask questions regarding homework and exam problems

12.2011-06.2012 **Teaching Assistant**
California State University, Fresno
Advisor: Nell Papavasiliou

- Assisted in the development of a new course for the engineering department geared towards teaching engineering students calculus from an applied engineering perspective
- Met weekly with course instructor to develop detailed lesson plans and activities to transform the calculus classroom into an interactive learning environment
- Implemented team-based learning strategies which allowed students to review and learn material before class time and participate in discussions and activities in class
- Presented select in class lectures and demonstrations that aided students in understanding calculus in terms of engineering

01.2009-01.2011 **Lyles College of Engineering Pathways Tutor**
California State University, Fresno
Advisor: Hernan Maldonado, Pathways Student Services Administrator

- Tutored students in a variety of engineering, math and science based courses
- Acted as a mentor and leader to students by assisting in resume and CV development and conducting mock interviews for and potential employment opportunities and graduate school
- Participated in ongoing teaching development seminars
- Tutored subjects included: *Calculus, Differential Equations, Linear Algebra, Physics, essay and report writing, Statics, Strength of Materials, Structures, Dynamics, Hydraulics, Environmental Engineering, Hydrology, Urban Storm*

water Management, Reinforced Concrete Design, Timber Design, Steel Design, Soil Engineering, Foundation Design

Work Experience

09.2019-present **Assistant Professor of Civil Engineering**
California State University, Fullerton

07.2012-07.2013 **Intern**
California Water Institute, Fresno
Supervisor: Sargeant Green, Water Management Specialist

- Worked with faculty and staff members in rotation with environmental justice representatives and other non-governmental organizations focused on water issues facing disadvantaged communities
- Assisted in data collection and analysis in preparation of education materials using the conduct of opinion surveys
- Research focused on the Central Valley aquifer system groundwater
- Compiled and analyzed data
- Compiled reports and answered questions pertaining to these reports

Leadership

09.2015-06.2016 **Strategic Planning Team Member**
Henry Samueli School of Engineering
University of California, Irvine

- Elected by the Dean Gregory Washington as the only graduate student in the Henry Samueli School of Engineering to serve of a committee of faculty and staff tasked with developing the strategic plan for the school for the next 10 years
- Attended bi-monthly meetings with the entire team for one academic year
- Jointly authored a report for the strategic plan

09.2015-06.2016 **Graduate Studies Committee Student Representative**
University of California, Irvine

- The Graduate Studies Committee (GSC) is comprised of Engineering Graduate Advisors, the Associate Dean for Research and Graduate Studies and the Graduate Student Representative
- Attend monthly committee meetings where school wide policies and procedures are discussed as they related to Engineering graduate students
- Provided student's perspective on various topics of discussion and stay aware of key and current graduate student issues
- Allowed voting rights for any procedures or policies that are voted on by the committee

01.2011-12.2011 **Student Project Manager**
California State University, Fresno

- Project Manager for a team of 8 students, elected by my peers
- Completed all structural design work for the building of a 289,000 square foot resort and hotel in Yosemite, California
- Collaborated with Rick Ransom and Associates in structural design work and industry mentorship throughout the project
- Team was awarded "Most Outstanding Project Report and Presentation" by a professional panel consisting of Civil Engineering faculty and members of industry

ABSTRACT OF THE DISSERTATION

Ground Motion Simulation Validation through Seismic Performance Assessment of Structural Systems

By

Huda Munjy

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2019

Professor Farzin Zareian, Chair

Recorded ground motions are traditionally used to represent future earthquakes for the seismic performance assessment and design of structures. However, there is a shortage of recorded ground motions that are able to represent a variety of possible scenarios (specifically large magnitude and short distance events). Structural engineers are hesitant in using simulated ground motions as surrogates for missing natural recordings, believing they are not equivalent to recorded ground motions in accurately estimating seismic demand in structures. This research recommends a validation test for simulated ground motions to increase the engineering community's confidence in using them in engineering applications.

This dissertation first introduces a novel statistical approach that can be used in current seismic performance assessment methods. Gaussian copulas are used to characterize the dependence structure of demand used for the assessment and calculation of loss in buildings due to earthquakes. Using Gaussian copulas is shown to increase the overall accuracy in seismic performance assessment methodology compared to the procedure that is currently being used, outlined by FEMA P-58 (2015). Following the proposal of a more accurate method to characterize

structural demand used in seismic performance assessment, a three step methodology, titled Vector-Based Intensity-measure Method (VBIM), for ground motion simulation validation at the structural response level is proposed. Three case study structures are used for the application of VBIM: 2- and 12-story special steel moment-resisting frame (SMRF) buildings and a two-span, cast-in-place concrete bridge. Results indicate that models of simulated and recorded ground motions that predict structural response based on waveform parameters are similar. The results of this study provide recommendations for ground motion simulators regarding the required accuracy of these key parameters in order for simulated ground motions to accurately predict structural response while also providing several steps of validation that show similarities between recorded and simulated ground motions.

CHAPTER 1: INTRODUCTION

1.1 PROBLEM STATEMENT

Simulated ground motions (GMs) are powerful tools in the behavioral analysis of structural systems. Simulated GMs may be applied in lieu of or as a compliment to recorded GMs during performance-based earthquake engineering research and design (Somerville et al., 2001). These simulations may be used as input for nonlinear dynamic analysis (NLDA) in structures (Bozorgnia and Bertero, 2004) or for generating earthquake intensity measures based off specified parameters (Somerville et al., 2001). In recent years, advancements in GM simulation methodologies have allowed for simulated GMs to be considered in scientific and engineering applications (Bradley et al., 2017). As the use of simulated GMs gains popularity, we must consider their impact on the accuracy of behavior prediction and design of structural systems in high seismicity areas when utilizing simulated GMs versus historical recorded GMs. Therefore, it is imperative that these simulation methodologies are not only validated to increase their use in engineering applications, but also in order to contribute to more reliable and accurate design of structures. This dissertation focuses on the validation of simulated ground motions at the seismic performance level through a vector of intensity measures that are statistically significant in predicting structural response of buildings and bridges.

In current structural and geotechnical engineering practice, the most common seismic inputs in design applications are real records from historical earthquakes, herein called recorded ground motions. Although these records represent actual ground motions that have occurred in the past, as a part of the procedure to utilize them as seismic inputs, they are often selected and scaled to match the target seismic hazard of the site in interest. This process of selection and scaling, called

Ground Motion Selection and Modification (GMSM), however, is the most widely accepted and used in engineering design, it does have limitations that need to be addressed. One specific limitation is the inability for GMSM techniques to accurately represent earthquake events with large magnitudes and small source to site distances as we do not have enough historical records with these features in general.

Simulated ground motions provide a viable alternative to be used in nonlinear dynamic analyses for the design of structures when there are not enough recorded ground motions available. Although design codes, such as ASCE Standard ASCE/SEI 7-10 (2010), allow for the use of simulated ground motions in seismic performance assessment, engineering practitioners are hesitant in choosing to use them over recorded ground motions because they are unsure of the similarities in the response of structures subject to simulated and recorded ground motions. In order to increase confidence in the engineering community towards the use of simulated ground motions in engineering applications, ground motions simulation validation efforts are needed. The concept of ground motion simulation validation dates back to the seminal works of Hartzell (1978) and Irikura (1978). Since then, many efforts have been made to develop and validate the use of simulated GMs in engineering applications (Zareian & Jones, 2010; Star et al, 2011; Seyhan and Stewart, 2014; Galasso et al, 2012; Galasso et al, 2013). Recently, many of the GM simulation validation efforts have been spearheaded by the Southern California Earthquake Center (SCEC) and focus on verification via the utilization of the Broadband Platform (BBP) (Dreger and Jordan, 2015).

This study provides a novel methodology for the validation of simulated ground motions at the seismic response level in terms of seismic performance assessment. Prior to directly tackling this validation effort, this study provides recommendations for increasing the accuracy of seismic

performance assessment used for performance based seismic design (PBSD) of structural systems through the implementation of more advanced statistical tools than what are currently recommended. There are three main components in PBSD: 1) ground motion hazard estimation, 2) structural response estimation, and 3) damage and loss estimation. To convey the results of PBSD to stakeholders, building performance is often measured through economic losses incurred due to seismic damage. In 2001, several organizations realized the great potential benefits of performance based seismic design (PBSD) research and implementation in the field of earthquake engineering—including the Federal Emergency Management Agency (FEMA), the Pacific Earthquake Engineering Research Center (PEER, a National Science Research Center) and the Applied Technology Council (ATC). In the same year, ATC began working with FEMA to develop Next-Generation Performance-Based Seismic Design Guidelines for New and Existing Building, which are under ATC-58/ATC-58-1. In 2012, FEMA P-58—*Seismic Performance Assessment of Buildings, Methodology, and Implementation*—was developed, which consists of a series of volumes with detailed recommendations and guidelines regarding seismic performance assessment via PBSD. Also included in this effort is the *Performance Assessment Calculation Tool* (PACT), which is software that practically implements the methodology, outlined in FEMA P-58 (2015).

In PBSA outlined by FEMA P-58, a joint lognormal distribution is used to represent the dependence structure of EDPs used for the quantification of damage and loss due to an earthquake. Several studies (Shome and Cornell, 1999; Song and Ellingwood, 1999; Shinozuka et al., 2000; Sasani and Kiureghian, 2001; Miranda and Aslani, 2004, Aslani and Miranda, 2005) assume that fitting a lognormal distribution to individual EDPs is statistically acceptable in most cases, but they did not test or address if the collection of EDPs' (EDP vector) true distribution is jointly

lognormal. Seismic performance assessment methodology, in general, relies heavily on the proper statistical representation of Engineering Demand Parameters (EDPs). However, current design guidelines assume that all EDPs have the same statistical dependence structure, i.e., a joint lognormal distribution is assumed. This assumption is restrictive and may lead to inefficiencies in the estimation of seismic losses in building structures. This study investigates the differences in estimated earthquake loss between generating realizations of the vector of peak floor accelerations (PFA) and the vector of maximum inter-story drift ratios (MaxIDR) using copulas to represent dependence and assuming a joint lognormal distribution. This is a preliminary step towards validating simulated ground motions at the seismic response level by first addressing potential areas of improvement in the methodology itself.

As the utilization of simulated ground motions increases in both the scientific and engineering communities, there is a growing need to validate these ground motions using a comprehensive set of intensity measures that are able to represent both the structures in consideration as well as the waveform of the ground motion at the structural response level. This dissertation builds upon the established aforementioned methods of ground motion simulation validation with the aim to validate simulated ground motion methodologies using a vector of intensity measures that are able to describe different elements of seismic performance, including waveform parameters and structure-dependent parameters. This is accomplished through a step by step procedural analysis that validates simulated ground motions at the seismic performance level, titled Vector Based Intensity-measure Method (VBIM) where significant intensity measures are used to predict future estimates of response. Key engineering demand parameters of different types of structures (i.e., bridge, buildings) via nonlinear dynamic analysis when subjected to bidirectional recorded and simulated ground motion components are determined. Through this analysis, a vector of diverse

intensity measures (i.e. structure-dependent, structure-independent) are tested to statistically determine the most significant proxies in predicting key engineering demand parameters of recorded motions which are then compared to ground motions simulated from several simulation methodologies. Furthermore, as vertical components of simulated ground motions have rarely been validated in previous studies (Bradley et al., 2017), this study utilizes both vertical and horizontal ground motion components in the validation methodology presented.

The results of this dissertation provide evidence of the similarities between the response of structures subject to recorded and simulated ground motions as well as predictions of future point estimates and mean estimates of response when specific intensity measures are accurately calculated and incorporated in the simulation methodologies. This can be used to further build confidence in the engineering community towards using simulated ground motions in engineering applications and also to inform ground motion modelers of which parameters need to be considered when designing and updating simulation methodologies in the future.

1.2 LITERATURE REVIEW

1.2.1 Seismic Performance Assessment Methodology

Multiple research efforts have been made to validate simulated ground motions. These efforts approach validation from different perspectives but mainly focus on validation through specific parameters or intensity measures that are highly associated with characterizing ground motions. This main goal of this dissertation is to provide a more holistic overview of simulated and recorded ground motions at multiple levels of response, starting from the actual waveform to the response of structures to each type of motion to measures of loss incurred. Therefore, before attempting to validate simulated ground motions at the seismic performance level, it is important to understand the theory and methodology recommended by design codes in seismic performance assessment. A

background on seismic loss assessment, specifically seismic loss assessment using Performance-Based Seismic Design, will be detailed here because the approach used in this dissertation to validate simulated ground motions is focused at the seismic performance and loss level. Therefore, it is necessary to understand the current guidelines regarding seismic performance, specifically from the PBSA perspective. The remainder of this section will outline FEMA P-58 (2015) methodology for the seismic performance assessment of structures.

One of the major goals of seismic performance assessment is to design buildings that, in the event of an earthquake, incur the least amount of damage and losses. One of the most capable tools in accomplishing this is Performance-Based Seismic Design (PBSD). PBSD takes into consideration the different types of earthquakes that could affect a building and determines the probability of experiencing certain levels of damage and loss. In PBSD, performance objectives are considered that relate to what the expected amount of damage may be as a result of an earthquake and what the specific consequences of that damage there may be. Current building codes, called prescriptive-based building codes, do not consider performance objectives, and instead design buildings based on a design-level earthquake where engineers design buildings based on specific criteria to meet the prescriptive building code standards. These buildings designed to code standards may adequately prevent casualties and injuries to human beings in the event of an earthquake but may not be designed to sustain extensive structural and non-structural damage. This leads to a large amount of possible downtime for buildings after earthquakes and repairs costs that are so high that it may be more economically feasible to demolish the building. Furthermore, there is little understanding of how well these typical prescriptive building codes actually live up to the level of seismic performance that the building was designed to meet. Generally, for many structures, these prescriptive-based building codes are sufficient in

minimizing casualties and injuries of human beings, but for critical structures that may need increased performance, PBSD should be implemented. PBSD allows for multiple stakeholders—including engineering designers, developers, and building owners—to come together and decide what level of building performance they are interested in, with of course still meeting typical building code standards.

In 2001, several organizations realized the great potential benefits of PBSD research and implementation in the field of earthquake engineering—including the Federal Emergency Management Agency (FEMA), the Pacific Earthquake Research Center (a National Science Foundation-funded Earthquake Engineering Research Center, PEER) and the Applied Technology Council (ATC). In the same year, ATC began working with FEMA to develop Next-Generation Performance-Based Seismic Design Guidelines for New and Existing Building, which are under ATC-58/ATC-58-1. In 2012, FEMA P-58—*Seismic Performance Assessment of Buildings, Methodology, and Implementation*—was developed which consists of a series of volumes with detailed recommendations and guidelines regarding seismic performance assessment via PBSD. Also included in this effort is *Performance Assessment Calculation Tool* (PACT), which is software that practically implements the methodology, outlined in FEMA P-58 (2015).

The first step in the seismic performance methodology outlined by FEMA is the assembly of the building performance model. This step includes defining both structural and non-structural components and building occupancy. Next, the earthquake hazard is defined for a specific site based on intensity and severity of the possible earthquakes in that region. After the earthquake hazard is defined, building response is analyzed in the form of engineering demand parameters that define the response. Some possible demand values include story drift, peak floor acceleration, and residual drift. Response can be determined via nonlinear response history analysis or for

regular, low- and mid- rise structures a simplified analysis using the equivalent lateral force procedure may be used. The next step is to develop collapse fragilities and then performance can be calculated in terms of a loss measure (e.g. casualties, downtime, and economic loss). Figure 1 summarizes these steps and also provides the metrics that are considered in each step.

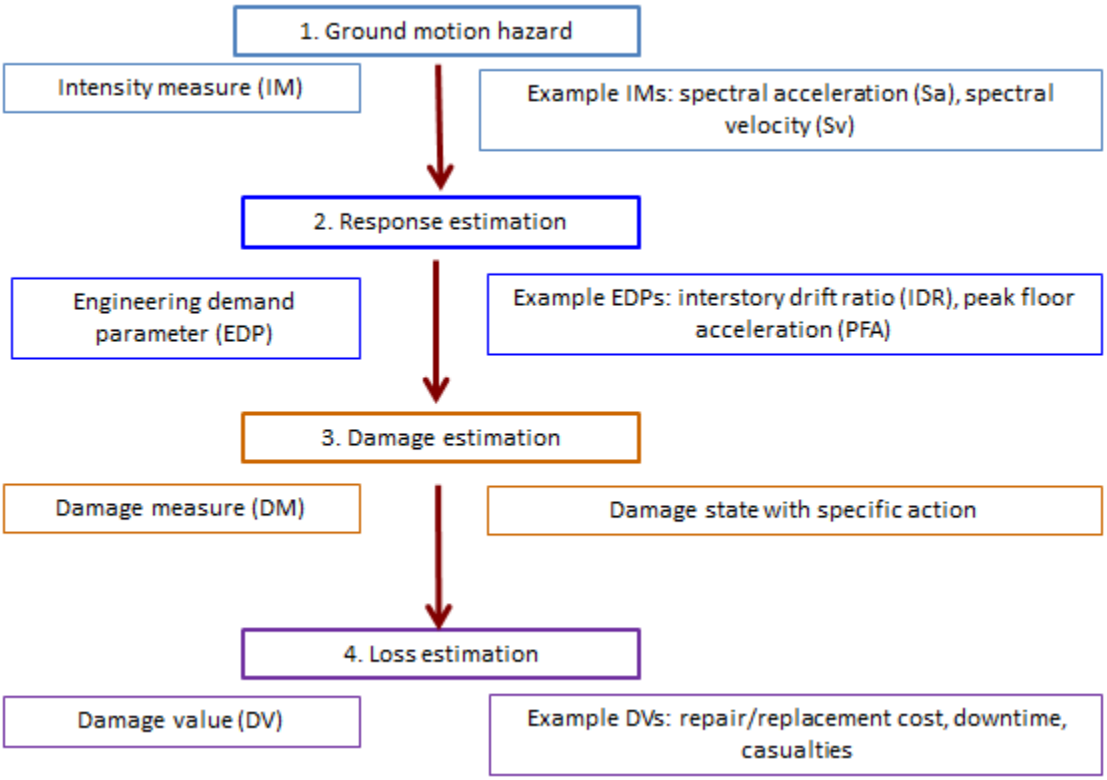


Figure 1.1: Outline of the steps for seismic performance assessment calculations as recommended by FEMA P-58 (2015)

In terms of the ground motion simulation validation procedure conducted in this dissertation, metrics at the ground motion hazard, response and loss levels are considered. The first part of this dissertation specifically focuses on the uncertainty in response estimation and loss estimation procedures, outlined by FEMA P-58 (2015), and proposes an updated statistical methodology to be implemented in the response estimation step in order to increase the accuracy and efficiency of the entire seismic performance assessment methodology. As outlined by FEMA P-58 (2015), simulated demand sets are determined through Monte Carlo procedure, which is used to generate a large number of demands from a small number of analyses by assuming that engineering demand parameters follow a joint lognormal distribution. The general procedure is to obtain the demand matrix, get median values for each parameter and a covariance matrix and then simulate a large number of demand vectors mathematically using a random number selection process and the median and covariance matrices. The required number of analyses depends on many factors but 11 analyses are mentioned in FEMAP-58 as an appropriate number. Yang (Yang et al, 2006; Yang et al, 2009) developed the algorithm used to generate the simulated demands (ATC, 2015). The peak absolute value of each EDP is assembled into a *1xn* vector. A joint lognormal distribution can be fully explained using the mean and covariance of the parameters, which are used in this analysis method. Cholesky decomposition is used as the matrix decomposition procedure. (1) and (2) are used to simulate EDPs according to FEMA recommendations where Z is the vector of natural log demand parameters, L represents the eigenvectors of the covariance demand matrix, D represents the eigenvalues of the covariance matrix and U is a matrix of uncorrelated standard normal variables with a mean of 0 and a covariance matrix which is the identity matrix.

$$Z = \lambda U + \mu \tag{1-1}$$

$$\lambda = LD \tag{1-2}$$

The main component of utilizing this method to generate large realizations of EDPs from a small number of initial data points is the assumption that EDPs possess a joint lognormal distribution. Although there are studies that confirm commonly considered EDPs, such as peak floor acceleration and inter-story drift ratio distributions over building height, can be represented by marginal lognormal distributions (Aslani & Miranda, 2005), there have been no studies to confirm that these EDPs possess a joint lognormal distribution. In order to address this assumption and the possible consequences of this assumption in terms of characterizing engineering demand and loss in buildings subject to earthquakes, a novel approach is proposed in the form of a statistical tool where several realizations of EDPs can be generated without having to assume any joint distribution between the parameters. This statistical tool replaces the need for the joint lognormal distribution assumption with the use of a statistical copula, which is used in statistics to provide a link between the marginal distributions of each single parameter and the joint distribution of all the parameters in the considered set.

Copulas are used to illustrate how marginal distributions can be linked together to describe a joint distribution. Therefore, given a copula, multivariate distributions can be determined even if variables follow different marginal distributions. According to Sklar's theorem (1959) the joint distribution function, $F_{XY}(x, y)$, can be expressed as a function of $F_x(x)$ and $F_y(y)$, which represent the marginal distributions of x and y, respectively (Nelson, 1999). The joint distribution function is represented by Equation (1-3), where $C(u,v)$ is the copula.

$$F_{XY}(x, y) = C(F_x(x), F_y(y)) \tag{1-3}$$

$$C(u, v) = F_{XY}(F_X^{-1}(u), F_Y^{-1}(v)) \tag{1-4}$$

$$C\big(F_X(x), F_Y(y)\big) = F(F_X^{-1}\big(F_X(x)\big), F_Y^{-1}(F_Y(y))) = F_{XY}(x, y) \quad (1\text{-}5)$$

When conducting a simulation using a copula distribution, the rank correlation of the data is used. A rank correlation is required when dealing with copulas because it measures the association based on the copula and not the marginal distributions of the variables. The common Pearson's correlation, measures the linear relationship between variables based on the marginal distribution and is not preserved by the copula. This means that two correlated variables that have the same copula can have different correlations. In contrast, the rank correlations are preserved. Kendall's tau, a rank correlation measure, is used. Tau is calculated for the data using R. Furthermore, the marginal distributions of the variables are required in order to develop the copula. A study by Aslani and Miranda (2005) confirms that the marginal distributions of EDPs follow a lognormal distribution.

1.2.2 Simulated Ground Motions

Traditional methods in earthquake engineering that are used to determine structural response and assess structural performance utilize recorded ground motions that have historically occurred. These recorded ground motions are usually selected and modified in order to match a specific target spectrum and the site under consideration. Although there exists a large and increasing database of recorded motions, there are not enough records to represent specific events, for example events with large magnitudes and short source to site distances. Furthermore, there are drawbacks that have been mentioned in previous literature (Baker, 2010) in scaling and modifying recorded motions that lead to concerns regarding how well they represent the recorded ground motions after modification. Simulated ground motions provide a code approved solution (ASCE Standard ASCE/SEI 7-10) that can describe the ground motion in terms of fault rupture, wave propagation and site response.

The first work in generating simulated ground motions was by P.C. Jennings (1968). Accelograms were generated to represent a variety of different earthquakes with magnitudes ranging from 5 to 8 through defining accelograms through sections of random processes. The next significant effort came from Hartzell (1978) which utilized aftershocks from large earthquake events as Green's functions. Boore (1983) proposed a simulation methodology that incorporates stochastic representations of source and path effects and Zeng (1994) proposed a fully deterministic method for simulation. Recent simulation methodologies that are being developed are either stochastic, deterministic or hybrid methodologies. There are several advantages and disadvantages of both stochastic and deterministic methods. For instance, stochastic methods are limited in lower frequency ranges and therefore, in these ranges deterministic approaches are favored. Deterministic methods, on the other hand, are said not to have enough complexity in higher frequency ranges to be accurate representations of recoded, or real, ground motions (Motazedian & Atkinson, 2005). The remainder of this section will detail the different simulation methodologies that are considered in this study.

1. Stochastic Finite-Fault (EXSIM)

   The Stochastic Finite-Fault (EXSIM) method was developed by Motazedian and Atkinson (2005) and utilizes dynamic corner frequency as a function of time and finite fault modeling of earthquake ground motions to provide a novel simulation methodology. Finite fault methodology corresponds to the division of a large fault into a certain number of subunits, with each subunit representing a point source. The ground motions from each subunit are summed and the ground motion along the entire fault is determined. This method is completely stochastic for all frequency ranges and does not incorporate deterministic elements.

2. Graves and Pitarka (GP)

The GP (Graves & Pitarka, 2014) simulation methodology is a hybrid approach which combines deterministic methods in the low frequency range and stochastic methods in the high frequency range. A single time history is produced from these two separate methods. This approach offers advantages over other hybrid simulation methodologies with the use of frequency-dependent non-linear site amplification factors. Overall this method reduces numerical computational requirements and increases the ability to incorporate site specific geologic information into the model.

3. Irikura Recipe

Irikura and Miyake (2011) developed a stochastic simulation methodology that is based on characterizing a source model for future earthquakes. This work identifies areas that correlate most with strong ground motions as asperities, which are defined as regions with large slip relative to the average slip of the entire rupture area. The validity of this method is tested by comparing recorded ground motions with ground motions simulated via IR.

4. Song

Song and Somerville (2010) develop a physics-based ground motion simulation methodology that is used to represent the physics involved in earthquake rupture processes. Geostatistics is implemented in order to develop the earthquake rupture process and to develop a source modeling tool to predict future earthquake events. The efficiency and accuracy of this method is tested using simulated and real rupture models to show that a large amount of the rupture process can be represented through this method.

1.2.3 Ground Motion Simulation Validation Efforts

Several GM simulation validation efforts have been spearheaded by the Southern California Earthquake Center (SCEC) and focus on verification via the utilization of the

Broadband Platform (BBP) (Dreger and Jordan, 2015). Several studies compare median spectral acceleration (S_a) values from different simulation methodologies with historical records and with values obtained from empirical ground motion prediction equations (Atkinson and Assatourians, 2015; Crempien and Archuleta, 2015; Graves and Pitarka, 2015; Olsen and Takedatsu, 2015; Star and Stewart, 2015). Along with Sa, other intensity measures (IMs) have also been used for the validation of simulated GMs (Burks and Baker, 2015; Razaein et al., 2015). Furthermore, studies have been conducted to propose a goodness-of-fit metric that measures the similarity between recorded and simulated GMs (Olsen and Mayhew, 2010).

As the utilization of simulated GMs increases in both the scientific and engineering communities, there is a growing need to validate these GMs using a comprehensive set of IMs that are able to represent both the structures in consideration as well as the waveform of the ground motion at the structural response level. This study builds upon the established aforementioned methods of GM simulation validation. Key engineering demand parameters (EDPs) of different types of structures (i.e., bridge, buildings) via NLDA when subjected to bidirectional recorded and simulated GM components are determined. Through this analysis, a set of diverse IMs (i.e. structure-dependent, structure-independent) are tested to statistically determine the most significant proxies in predicting key EDPs of recorded motions which are then compared to GMs simulated from several simulation methodologies. Furthermore, as vertical components of simulated GMs have rarely been validated in previous studies (Bradley et al., 2017), this study utilizes both vertical and horizontal GM components in the validation methodology presented.

This study not only compares parameters related to S_a as proxies for validation with values from simulated, recorded and empirically predicted GMs, but also parameters of GMs that have physical meanings and that are not structure-dependent. Four parameters, detailed by Razaeian and

Der Kiureghian (2010), that contribute to describing the time modulation and evolutionary frequency content of nonstationary GM models are considered as proxies along with S_a. These four parameters—Arias intensity (I_a), effective duration of the motion ($D_{5\text{-}95}$), filter frequency (ω_{mid}) at the time in which 45% of I_a is reached and the rate of change of the filter frequency with time (ω')—have been proposed as validation metrics and proven to define ground motion waveforms, but have not been tested for significance in predicting the response of structures, which is a key element in seismic performance assessment methodology (Rezaeian et al, 2015).

The aim of this study is to validate simulated ground motion methodologies using a vector of intensity measures that are able to describe different elements of seismic performance, including waveform parameters and structure-dependent parameters. This is accomplished through a step by step procedural analysis that validates simulated ground motions at the seismic performance level, titled Vector Based Intensity-measure Method (VBIM) where significant intensity measures are used to predict future estimates of response. Multiple regression analyses are carried out to statistically determine which parameters contribute to predicting response of structures subject to 1994 Mw 6.7 Northridge recorded earthquake and its simulated counterparts considering four simulation methodologies available via Southern California Earthquake Center (SCEC) Broadband Platform (BBP): 1) stochastic finite-fault (EXSIM) (Motazedian and Atkinson, 2005; Atkinson et al., 2009; Boore, 2009), 2) hybrid simulations by Graves and Pitarka (GP) (Graves and Pitarka, 2014), 3) Irikura recipe (IR) (Irikura and Miyake, 2011), and 4) Song (Song & Somerville, 2010).

1.2.3 Metrics Used to Validate Simulated Ground Motions

One challenging aspect of validating simulated ground motions is choosing a metric that is able to accurately define ground motion waveforms and that is easily interpretable to multiple

stakeholders. A study conducted by Rezaeian et al (2015) highlights several advantages in choosing metrics for the validation of simulated ground motions that are able to capture the entire evolution of ground motion intensity and frequency content, as opposed to intensity measures that only define specific points in the waveform. One such advantage is that by capturing the entire evolution of intensity and frequency in these metrics, structural response is expected to also be accurately characterized. As this study focuses on validating simulated ground motions at the structural response level, four metrics that were previously studied by Rezaeian et al (2015) are used for validation. We consider two intensity based metrics, Arias intensity (I_a) and significant duration of motion ($D_{5\text{-}95\%}$), and two frequency based metrics, predominant frequency at mid-duration (ω_{mid}) and slope of predominant frequency at mid duration (ω'). Both I_a and $D_{5\text{-}95\%}$ are representative of energy accumulation and have been shown to affect the response of structural systems (Dashti et al, 2010; Ghayoomi and Dashti, 2014). Therefore, it is worthwhile to investigate how these metrics affect the response of structures subject to not only historical earthquakes but their simulated counterparts as well. In addition to these waveform parameters, spectral acceleration (S_a) is used as an additional metric to show similarities in the response of structures subject to recorded and simulated ground motions. S_a has been used in several other validation efforts, mentioned previously, and therefore is included in this study as it has been proven to be a useful metric for ground motion simulation validation.

1.3 HYPOTHESES AND RESEARCH PLAN

In this dissertation I hypothesize that simulated ground motions can be accurate representations of recorded ground motions if certain significant intensity measures are accurately incorporated into the simulation methodology. Furthermore, I hypothesize that, if this incorporation is successfully accomplished, then structural response and loss during for a structure

designed based on simulated ground motions will be within a reasonable margin of error of structural response and loss during for a structure designed based on recorded ground motions. With this ultimate goal, the objectives of this research are as follows:

1. Identify intensity measures (both waveform-dependent and structure-dependent) that are significant in predicting the response of structures subject to recorded ground motions
2. Statistically compare the models that are developed for the prediction of structural response based on significant intensity measures for simulated and recorded earthquakes
3. Provide a means to measure how differences in significant intensity measures between simulated and recorded ground motions correspond to differences in response of structures subject to simulated versus recorded ground motions
4. Extend this measure to understand how differences in significant intensity measures between simulated and recorded ground motions correspond to differences in loss of structures subject to simulated versus recorded ground motions

Objectives 1 through 3 are accomplished within the scope of this dissertation and will be detailed through the remaining chapters. Further research is required to complete task 4, but the work will still be addressed here as it is the next step that in this process and will contribute to a more holistic outlook on the validation of simulated ground motions at multiple levels of seismic performance assessment.

1.4 REFERENCES

American Society of Civil Engineering (ASCE) (2010). Minimum design loads for buildings and other structures (7–10), Standards ASCE/SEI 7–10.

ATC - Applied Technology Council, FEMA P-58 [2015b] *Next-generation Seismic Performance Assessment for Buildings, Volume 2 – Implementation Guide*, Federal Emergency Management Agency, Washington, D.C., 2015.

Atkinson, G., Assatourians, K., Cheadle, B., & Greig, W. (2015). Ground motions from three recent earthquakes in western Alberta and northeastern British Columbia and their implications for induced-seismicity hazard in eastern regions. *Seismological Research Letters*, *86*(3), 1022-1031.

Baker, J. W. (2010). Conditional mean spectrum: Tool for ground-motion selection. *Journal of Structural Engineering*, *137*(3), 322-331.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America*, *73*(6A), 1865-1894.

Bozorgnia, Y., & Bertero, V. V. (2004). *Earthquake engineering: from engineering seismology to performance-based engineering*. CRC press.

Bradley, B. A., Razafindrakoto, H. N., & Polak, V. (2017). Ground-motion observations from the 14 November 2016 M w 7.8 Kaikoura, New Zealand, earthquake and insights from broadband simulations. *Seismological Research Letters*, *88*(3), 740-756.

Burks, L. S., & Baker, J. W. (2016). A predictive model for fling-step in near-fault ground motions based on recordings and simulations. *Soil Dynamics and Earthquake Engineering*, *80*, 119-126.Razaein et al., 2015

Crempien, J., & Archuleta, R. (2015). UCSB method for broadband ground motion from kinematic simulations of earthquakes. *Seismological Research Letters, 86*(1).

Dreger, D., Beroza, G., Day, S., Goulet, C., Spudich, P., Stewart, J., & Jordan, T. (2015). Evaluation of SCEC broadband platform phase 1 PSA ground motion simulation results. *Seismological Research Lett*ers *86*(1).

FEMA (2015) Next-Generation Methodology for Seismic Performance Assessment of Buildings, prepared by the Applied Technology Council for the Federal Emergency Management Agency, Report No. FEMA P-58, Washington, D.C.

Galasso, C., Zareian, F., Iervolino, I., & Graves, R. W. (2012). Validation of ground-motion simulations for historical events using SDoF systems. *Bulletin of the Seismological Society of America*, *102*(6), 2727-2740.

Galasso, C., Zhong, P., Zareian, F., Iervolino, I., & Graves, R. W. (2013). Validation of ground-motion simulations for historical events using MDoF systems. *Earthquake Engineering & Structural Dynamics*, *42*(9), 1395-1412.

Graves, R., & Pitarka, A. (2015). Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters*, *86*(1), 75-80.

Hartzell, S. H. (1978). Earthquake aftershocks as Green's functions. *Geophysical Research Letters*, *5*(1), 1-4.

Irikura, K. (1983). Semi-empirical estimation of strong ground motions during large earthquakes.

Irikura, K., & Miyake, H. (2011). Recipe for predicting strong ground motion from crustal earthquake scenarios. *Pure and Applied Geophysics*, *168*(1-2), 85-104.

Jennings, P. C., Housner, G. W., & Tsai, N. C. (1968). Simulated earthquake motions.

Jones, P., & Zareian, F. (2010). Relative safety of high-rise and low-rise steel moment-resisting frames in Los Angeles. *The Structural Design of Tall and Special Buildings*, *19*(1-2), 183-196.Star et al, 2011;

Miranda, E., Aslani, H., & Taghavi, S. (2005). Assessment of seismic performance in terms of economic losses. In *Proceedings, International Workshop on Performance-Based Seismic Design: Concepts and Implementation* (Vol. 28, pp. 149-160). Berkeley: Pacific Earthquake Engineering Research (PEER) Center, University of California.

Motazedian, D., & Atkinson, G. M. (2005). Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America*, *95*(3), 995-1010.

Olsen, K. B., & Mayhew, J. E. (2010). Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake. *Seismological Research Letters*, *81*(5), 715-723.

Rezaeian, S., & Der Kiureghian, A. (2010). Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthquake Engineering & Structural Dynamics*, *39*(10), 1155-1180.

Rezaeian, S., Zhong, P., Hartzell, S., & Zareian, F. (2015). Validation of simulated earthquake ground motions based on evolution of intensity and frequency content. *Bulletin of the Seismological Society of America*, *105*(6), 3036-3049.

Seyhan, E., & Stewart, J. P. (2014). Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. *Earthquake Spectra*, *30*(3), 1241-1256.

Somerville, P., Collins, N., Abrahamson, N., Graves, R., & Saikia, C. (2001). Ground motion attenuation relations for the central and eastern United States. *US Geological Survey, Award 99HQGR0098, final report.*

Song, S. G., & Somerville, P. (2010). Physics-based earthquake source characterization and modeling with geostatistics. *Bulletin of the Seismological Society of America*, *100*(2), 482-496.

Star, L. M., Givens, M. J., Nigbor, R. L., & Stewart, J. P. (2015). Field-testing of structure on shallow foundation to evaluate soil-structure interaction effects. *Earthquake spectra*, *31*(4), 2511-2534.

Yang, T. Y., Moehle, J., Stojadinovic, B., & Der Kiureghian, A. (2006, April). An application of PEER performance-based earthquake engineering methodology. In *Proceedings*.

Yang TY, Moehle JP, Stojadinovic B, Der Kiureghian A (2009) Seismic performance evaluation of facilities: methodology and implementation. J Struct Eng-ASCE 135(10):1146–115.

CHAPTER 2: CHARACTERIZING DEPENDENCE OF ENGINEERING DEMAND PARAMETERS USED IN SEISMIC LOSS ESTIMATION USING STATISTICAL METHODS

2.1 INTRODUCTION

Two novel statistical approaches for modeling the dependence structure of engineering demand parameters (EDPs) that are used in seismic performance assessment procedures is proposed and evaluated against recommendations provided by FEMA P-58 (2015). These current provisions outlined by FEMA P-58 (2015) utilize Monte Carlo simulation is to generate realizations of engineering demand parameters. EDPs are simulated by first taking the log of the original data and then calculating the mean of each random variable and the covariance matrix of the logged data. Next, standard normal random numbers are generated and are combined with the covariance matrix of the data to generate a certain number of realizations. In the last step, the means of each random variable are added to the simulated data.

The required assumption in this procedure is that the natural logs of EDPs follow a joint normal distribution because normally distributed random numbers are generated and used for this

simulation procedure. Furthermore, Cholesky decomposition (CD) is used; according to Straka et al (2013) for random variables with correlated elements, eigenvalue decomposition (ED) is the preferred method for matrix decomposition.

The first proposed approaches utilizes copula theory, a statistical tool that is used in many applications. For example, copulas are used in the field of business and financial risk management to measure risk across a diverse range of areas in finance (Habiboellah, 2007). Using copula theory in this research was chosen as it poses several advantages to traditional approaches used to generate engineering demand parameters. These advantages include: 1) engineering demand parameters can all have different marginal distributions, 2) no assumption is required for the joint distribution of the variables, 3) there are several widely available copula families (i.e. distributions) to choose from that represent the link between the variables which make copulas applicable to a wide variety of problems.

The copula distribution represents a link between marginal distributions of variables and the dependencies of these variables that make up the joint distribution (Nelson, 1999). Using a copula distribution allows for the simulation of random variables that may each follow a different marginal distribution. Aslani and Miranda (2005) found that PFA and IDR are marginally lognormally distributed; therefore, PFA and IDR are simulated using a copula distribution and lognormally distributed marginal distributions. Furthermore, when using a copula distribution, rank correlation is used as opposed to linear correlation. A rank correlation is required when dealing with copulas because it measures the association based on the copula and not the marginal distributions of the variables (Nelson, 1999). Linear correlation measures the linear relationship between variables based on the marginal distribution and is not preserved by the copula. This means that two correlated variables that have the same copula can have different correlations; in

contrast, rank correlations are preserved. Kendall's tau, a rank correlation measure, is used in this study. R statistical analysis software (R Core Team, 2013) and MATLAB (MATLAB, 2016) are used to simulate EDPs using the copula distribution. MATLAB is used to generate random values for the copula distribution which links the marginal distributions with the joint distribution of the EDPs using the rank correlation of the variables. The inverse transform sampling method is used to generate random numbers following the same distribution as the input data using the inverse cumulative distribution function (CDF). Inverse transform sampling simulates random numbers with the same distribution as the input data by obtaining the probability distribution functions (PDFs) of the random variables and creating a CDF from this PDF (a Riemann sum can be used to accomplish this) and inverting this CDF (Devroye, 1986). An arbitrary value, *e*, can be chosen with a uniform distribution and the inverted value of this is equal to *y*; *invCDF(e)=y*. This value of *y* is our desired random number from our original random variable with the desired PDF. A kernel smoothing function is used to estimate the inverse CDF for each random variable. Therefore, a copula can be used to represent the link between the marginal distributions of EDPs, which were confirmed by Aslani and Miranda (2005) to be lognormal, and the joint distribution of the variables.

Along with copula theory, factor analysis is applied to the generation of realizations of EDPs. Factor analysis is a dimension reduction tool where the first principal component is the dimension that is able to explain the most variation within the data, and with each successive component explaining less variation. Factor analysis takes *n* observed data points on a vector *x* of *p* random variables and is able to reduce the dimension of the dataset from the original *p* to a new, usually much smaller number, *q* linear combinations, called factors. These linear combinations are in the form $a_1'x$, $a_2'x \ldots, a_q'x$, where $a_1, a_2, \ldots, a_q$ are the eigenvectors of the covariance matrix of the

data that correspond to the q largest eigenvalues (Jolliffe, 2011). One of the main advantages of using factor analysis as a statistical tool for generating realizations of EDPs is its usefulness in reducing the number of variables to a much smaller number, namely components. This is especially useful for tall buildings (considering PFA and IDR, a 20-story building would have 41 EDPs). Furthermore, factor analysis simulates the dependence structure of EDPs using Eigen decomposition as opposed to Cholesky decomposition, which is currently implemented in FEMA P-58 methodology. Cholesky decomposition is the preferred matrix decomposition for uncorrelated random variables but studies show that for correlated random variables, Eigen decomposition provides more accurate estimates of mean and covariance structure for simulations (Straka et al, 2013). Therefore, factor analysis may be able to provide more accurate realizations of EDPs compared with other methods. Unlike copula theory and FEMA P-58 recommended approach for the simulation of EPDs, factor analysis has certain restrictions that may make it less favorable for the process of nonlinear response history analysis in seismic performance assessment. One particular drawback is that recommendations for using factor analysis strongly suggest the use of a full rank covariance matrix for the data. This means that the number of EDP variables should not exceed the number of actual data points available. Nonetheless, factor analysis is studied here for the main goal of providing a comparison of how EDP realizations generated with full rank covariance matrices compare to EDP realizations when we have more variables that actual data inputs.

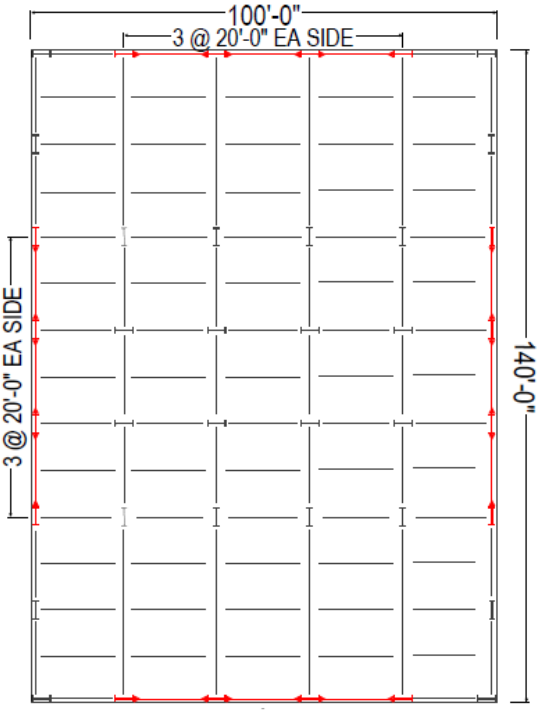
2.2 DESCRIPTION OF CASE STUDY STRUCTURES AND GROUND MOTION SELECTION

Four special steel moment resisting frames (SMRFs) are utilized as case study structures which include 2-, 4-, 8- and 12- story buildings with corresponding first-mode building periods of

0.92, 1.61, 2.28 and 3.10 seconds, respectively. These buildings are designed based on ASCE/SEI 7-02 (ASCE, 2005) and ANSI/AISC 341-05 (AISC, 2005) for a site in downtown Los Angeles, California with typical soil site class D. Plan and elevation views of the buildings are shown in Figure 1. Static pushover curves, shown in Figure 2, are generated for the buildings to show the general load-deflection relationship.

Ground motion selection and scaling (GMSM) is done for a total of 200 ground motions per case study building based on the Conditional Mean Spectrum (CMS) method (Baker, 2010). Disaggregation for all buildings is accomplished using Open Source Seismic Hazard Analysis (Field et al, 2005). GMSM is performed for design basis earthquake (DBE) level ground motions considering 10% probability of exceedance in 50 years and 50% probability of exceedance in 50 years. The ground motions are selected from the Next Generation of Attenuation Relationships (PEER-NGA) database, which does not include records of event foreshocks or aftershocks. 100 ground motions per hazard level (100 for 10% in 50 years and 100 for 50% in 50 years) per building are selected and scaled to match the target conditional mean spectra for the location of the buildings (all buildings are modeled at the same location).

Through nonlinear dynamic analyses (NLDA), engineering demand parameters (EDPs) of peak floor acceleration (PFA) and maximum inter-story drift ratio (IDR) at all story levels are obtained. The EDPs generated corresponding to these 100 selected and scaled ground motions per hazard level per case study structure are referred to as population EDPs, as they are directly generated via NLDA and no simulations based on assumptions of EDP dependence are conducted. These EDP populations are used as the baselines for comparisons with the EDPs generated under copula and joint lognormal distribution assumptions.



SMRF Beam and Column Schedule

Beam		Column	
Label	Size	Label	Size
SMF BM1	W21x57	SMF C1	W24x31
SMF BM2	W21x73	SMF C2	W24x62
SMF BM3	W16x31	SMF C3	W24x33
SMF BM4	W30x132	SMF C4	W24x103
SMF BM5	W21x68	SMF C8	W24x84
SMF BM6	W24x84	SMF C9	W24x146
SMF BM7	W27x94	SMF C10	W24x162
SMF BM8	W30x116	SMF C11	W24x207
SMF BM9	W30x108		

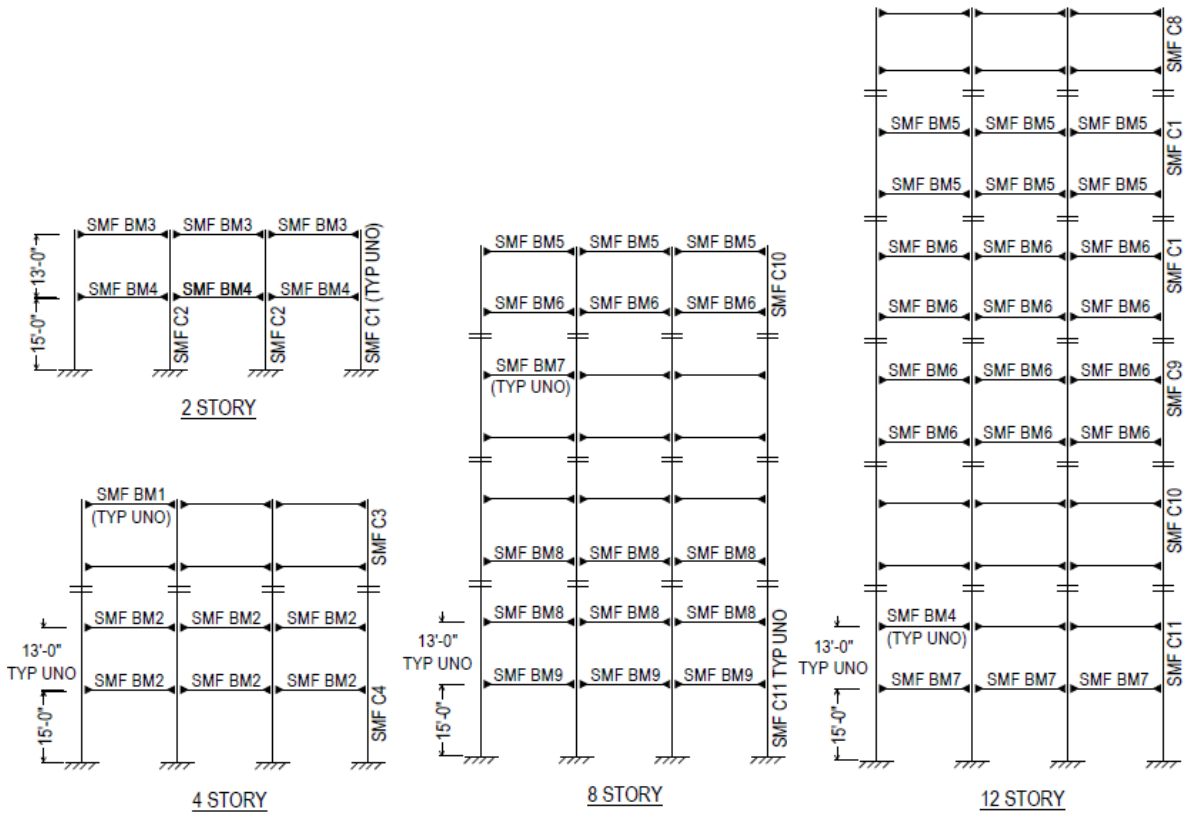


Figure 2.1: Plan and elevation views for 4- and 8-story case study buildings

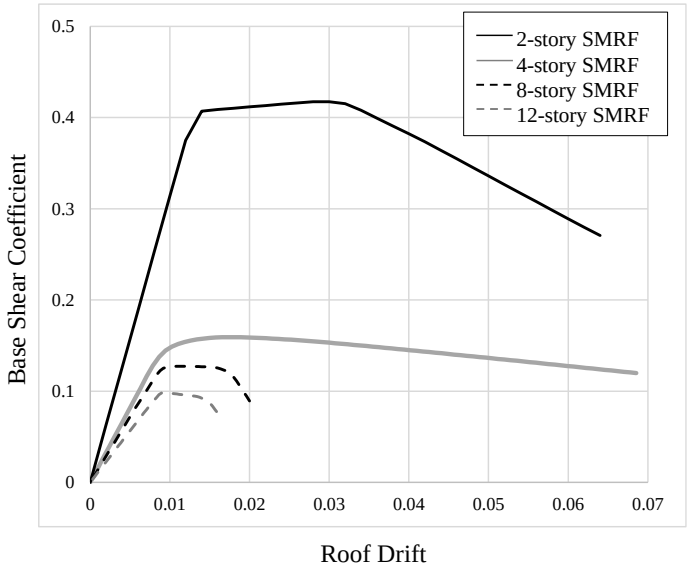


Figure 2.2: Static pushover curves for 4- and 8-story SMRF case study buildings

The first part of this study requires the selection of ground motions to be used for the seismic performance assessment of the structures. Ground Motion Selection and Modification (GMSM) is performed as per Jayaram and Baker's (2011) study which provides an algorithm for matching target mean spectrums of ground motion records. This GMSM method is incorporates the variance of the target response spectrum in a computationally efficient and theoretically sound manner, which few other methods do. Multiple response spectra are generated from the target distribution and the individual ground motions are selected with spectras that match the target spectrum. FEMA P-58 Volume 1 (2015) recommends the use of 7 to 11 analyses, or ground motions, to obtain valid estimates of median response. More analyses are recommended when the geomean spectral shape of scaled motions does not accurately match the shape of the target spectrum. It is suggested that 11 analyses be used for ground motions that are selected with no consideration to the spectral shape (Huang et al, 2011). There is little reasoning provided as to why 11 ground motions are sufficient in producing unbiased estimates of median structural response so for this study 100 ground motions are selected and modified in order to provide a wide range of possible ground motions and resulting EDPs from these 100 ground motions. These 100 ground motions are utilized as the population ground motions, or in other words as all of the possible scenarios of ground motions that could hit the buildings. From this considered population of 100 ground motions, a number of samples are selected for analyses to understand how EDPs generated using FEMA recommendations, copula theory, and factor analysis differ. Two hazard levels are considered: 10% exceedance in 50 years and 50% exceedance in 50 years. Therefore, for each hazard level, 100 ground motions were selected and modified providing two separate ground

motion populations based on hazard level. Every analysis detailed herein for this section is done for both hazard levels and compared.

From the selected and modified population of 100 ground motions, 11 ground motions are randomly selected and values of PFA and IDR are generated via nonlinear response history analysis for both the 4- and 8-story building. The reasoning behind choosing 11 ground motions is because of the FEMA P-58 recommendation that state 7-11 ground motions are sufficient for determining the EDPs used in seismic performance assessment. The higher number from this range was used to select the sample ground motions. After the random selection of the 11 ground motions, these ground motions were subject to the two buildings and via nonlinear response history analysis, values of PFA and IDR were determined. All building design and analyses including the generation of these EDPs were conducted using the Open System for Earthquake Engineering Simulation (OpenSees) software (http://opensees.berkeley.edu/).

2.3 PERFORMANCE BASED SEISMIC DESIGN METHODOLOGY

The accurate assessment of loss is essential in minimizing earthquake associated risks. The performance based seismic design (PBSD) approach aims to accurately assess loss through the design of structures that meet seismic performance objectives, which can be classified in terms of dollars, death, or downtime. In PBSD, successive numerical integrations of conditional probabilities in terms of site hazard, structural response, damage response, and loss response are combined where uncertainties are propagated in each step.

There are three main components in PBSD: 1) ground motion hazard estimation, 2) structural response estimation, and 3) damage and loss estimation. In order to convey the results of PBSD to stakeholders, building performance is often measured through economic losses incurred due to damage. In 2001, several organizations realized the great potential benefits of

performance based seismic design (PBSD) research and implementation in the field of earthquake engineering—including the Federal Emergency Management Agency (FEMA), the Pacific Earthquake Engineering Research Center (PEER, a National Science Research Center) and the Applied Technology Council (ATC). In the same year, ATC began working with FEMA to develop Next-Generation Performance-Based Seismic Design Guidelines for New and Existing Building, which are under ATC-58/ATC-58-1. In 2012, FEMA P-58—*Seismic Performance Assessment of Buildings, Methodology, and Implementation*—was developed which consists of a series of volumes with detailed recommendations and guidelines regarding seismic performance assessment via PBSD. Also included in this effort is *Performance Assessment Calculation Tool* (PACT), which is software that practically implements the methodology, outlined in FEMA P-58 (2015). Baker et al (2016) utilize the 2010-2011 Canterbury earthquake sequence and resulting extensive data sets of structural damage to present and evaluate FEMA P-58 (2015) guidelines. Although an extensive analysis of results is not provided, there are noted differences between results from FEMA P-58 and from the actual loss incurred to the structural systems during the Canterbury earthquake.

In PBSA outlined by FEMA P-58 (2015), a joint lognormal distribution is used to represent the dependence structure of EDPs used for the quantification of damage and loss due to an earthquake. Several studies (Shome and Cornell, 1999; Song and Ellingwood, 1999; Shinozuka et al, 2000; Sasani and Kiureghian, 2001; Miranda and Aslani, 2003, Aslani and Miranda, 2005) assume that fitting a lognormal distribution to individual EDPs is statistically acceptable in most cases, but do not test or address if the collection of EDPs' (EDP vector) true distribution is jointly lognormal. Peak floor acceleration (PFA) and maximum interstory drift ratio (maxIDR) are commonly used in PBSD, as they are representative of acceleration sensitive and drift sensitive

components, respectively, and therefore provide a complete overview of building response. Aslani and Miranda (2005) found that peak floor acceleration (PFA) did not fit a lognormal distribution as well as other EDPs but was still within an acceptable range when collapse cases were removed. The quantification of earthquake induced damage and subsequent loss is highly influenced by the computation and modelling of EDPs. Goda (2010) studied modeling of two demand parameters, namely maximum interstory drift ratio (MaxIDR) and residual interstory drift ratio (ResIDR), through single degree of freedom structures. Goda and Tesfamariam (2015) investigated the joint probabilistic modeling of multiple EDPs using copulas, specifically for MaxIDR and ResIDR, and developed a multivariate seismic demand model that relates economic loss from earthquakes with specific cases including non-collapse, collapse, and demolition.

Seismic performance assessment methodology, in general, relies heavily on the proper statistical representation of engineering demand parameters (EDPs), however, current design guidelines assume that all EDPs have the same statistical dependence structure, i.e. a joint lognormal distribution is assumed. This assumption is restrictive and may lead to inefficiencies in the estimation of seismic losses in building structures. This study investigates the differences in earthquake loss between generating realizations of peak floor acceleration (PFA) and maximum interstory drift ratio (MaxIDR) using copulas to represent dependence and assuming a joint lognormal distribution. These two EDPs are chosen because of their respective influence on acceleration and drift, therefore encompassing a demand vector capable of presenting a holistic overview of response. Furthermore, a Kolmogorov-Smirnov test is conducted to determine the optimal number of initial ground motions required for accurate assessments of response when EDPs are generated using copulas and also assuming a joint lognormal distribution. Copulas have been used in a wide variety of applications, especially in the field of financial risk management to

measure risk across a diverse range of areas in finance (Habiboellah, 2007), as they provide a convenient way to model the dependence between multi-variate data (Goda and Tesfamariam, 2015; Genest and Favre, 2007). Both of these methods for generating demand realizations are generating seismic loss are applied to 4 story special steel moment resisting frame buildings located in Los Angeles, California, a high seismicity region with expected earthquake risk of hundreds of millions of dollars per year (Olshanksky and Yueming, 2001). The contributions of this study are (1) the application and assessment of two separate methodologies for modeling the dependence between multivariate data (i.e. copulas and joint lognormal distribution) for generating values of economic loss in buildings due to earthquakes, and (2) recommendations regarding the number of ground motions required in seismic response assessment calculations to produce accurate measures of building response.

The performance based earthquake engineering (PBEE) methodology is used to probabilistically assess the vulnerability of structural systems during a seismic event. PBEE allows for multiple stakeholders—including engineering designers, developers, and building owners—to come together and decide what level of building performance they are interested in, with of course still meeting typical building code standards. The typical methodology of PBSD includes hazard analysis, structural analysis and damage-loss analysis. In the structural analysis portion, several realizations of demand parameters are generated based on a smaller number of initial inputs. In mathematical terms, performance based earthquake engineering methods can be represented in the form of a triple integral that is based on the total probability theorem (Moehle and Deierlein, 2004):

$$v(DV) = \int\int\int G\langle DV|DM\rangle|dG\langle DM|EDP\rangle|dG\langle EDP|IM\rangle|d\lambda(IM) \quad (2\text{-}1)$$

where *IM* is the intensity measure determined from hazard analysis, *EDP* is the engineering demand parameter obtained through structural analysis, *DM* corresponds to the damage measure

from damage analysis, *DV* is the decision variable from the loss analysis, and $\lambda(IM)$ is the mean annual rate of exceedance from probabilistic seismic hazard analysis procedures. This mathematical representation is complex and encompasses several layers of variability. In this study, we explore the variability that exists directly after the structural analysis, in the processes that can be used to generate realizations of EDPs, and also the variability in loss that exists based on these processes. Spectral acceleration at the fundamental period of vibration is used as a single variable IM, while PFA and MaxIDR are used as the multi-variate EDPs.

As outlined by FEMA P-58 (2015), simulated demand sets are determined through Monte Carlo procedure, which is used to generate a large number of demands from a small number of analyses by assuming that EDPs follow a joint lognormal distribution. The general procedure is to obtain the demand matrix, determine the median values for each parameter and a covariance matrix, and then simulate a large number of demand vectors mathematically using a random number selection process and the median and covariance matrices. The required number of analyses depends on many factors, but 11 analyses are mentioned in FEMAP-58 as an appropriate number. Yang (Yang et al, 2006; Yang et al, 2009) developed the algorithm used to generate the simulated demands (ATC, 2015). The peak absolute value of each EDP is assembled into a *1xn* vector. A joint lognormal distribution can be fully explained using the mean and covariance of the parameters, which are used in this analysis method. Cholesky decomposition is used as the matrix decomposition procedure. Equations (2-2) and (2-3) are used to simulate EDPs according to FEMA recommendations where *Z* is the vector of natural log demand parameters, *L* represents the eigenvectors of the covariance demand matrix, *D* represents the eigenvalues of the covariance matrix and *U* is a matrix of uncorrelated standard normal variables with a mean of 0 and a covariance matrix which is the identity matrix.

$$Z = \lambda U + \mu \tag{2-2}$$

$$\lambda = LD \tag{2-3}$$

2.4 PROPOSED STATISTICAL METHODS FOR MODELING DEPENDENCE OF EDPS

2.4.1 Modeling dependence using copulas

Copulas are mathematical tools that are able to capture the dependence structure of multivariate data. They have several applications and have been widely applied in the fields of finance, insurance, reliability theory, and more recently, in the field of hydrology (Jaworski et al, 2009; Genst and Favre, 2007).

Consider random variables X and Y whose distribution functions are $F(x) = P[X \leq x]$and, $G(x) = P[Y \leq y]$, respectively. The joint distribution of X and Y is therefore $H(x, y) = [X < x, Y < y]$. For all real numbers (x, y) we now have three other numbers that can be associated: $F(x)$, $G(y)$, and $H(x, y)$, with all of these numbers lying in the interval $[0,1]$. Therefore, each ordered pair $(F(x), G(y))$ corresponds to a number for the joint distribution $H(x, y)$ that lies within $[0,1]$ and that this link between the ordered pair and the joint distribution is called a copula (Nelsen, 2007) and according to Sklar's theorem (Sklar, 1973), this copula is represented by *C* in (2-4).

$$H(x, y) = C\big(F(x), G(y)\big) \tag{2-4}$$

Using Sklar's theorem, the joint distribution can be modelled using the marginal distributions of the random variables and the copula, without having any information regarding the joint distribution of the variables. This allows for greater flexibility when modeling the dependence structure of random multivariate data. For the dependence structure of X and Y, an empirical copula is used, represented by (2-5) where R_i and S_i are the rank of X_i and Y_i in ascending order, respectively and u and v represent $F(x)$ and $F(y)$, respectively (Genest and Favre, 2007).

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \quad (2\text{-}5)$$

One of the main difficulties in using copulas to model the dependence structure of variables is choosing the parametric family or distribution of the copula that best represents the link between the data. A Gaussian copula is used in this study as it is not only a common choice for multivariate data but also because previous literature has shown that a Gaussian copula is able to capture characteristics of the dependence structure of select maximum interstory drift ratio and peak floor acceleration well (Goda and Tesfamariam, 2015).

2.4.2 Modeling dependence using factor analysis

Factor Analysis is a statistical procedure that can be used to detect interrelationships between variables. This procedure has a number of applications that revolve around data reduction and identifying links among different factors used to assess data. The data is effectively reduced by determining a smaller number of unobserved factors from a larger number of variables. The two assumptions of factor analysis are that the error terms are independent of each other and the factors are independent of each other. Factor Analysis therefore aims to explain correlations among a large number of observed variables by describing these variables with a smaller number of unobserved 'factors'.

2.5 EVALUATION OF ENGINEERING DEMAND PARAMETER DISTRIBUTION

Assuming lognormality of EDPs is common in performance based earthquake engineering research. This assumption is tested through the use of three statistical tests that assess the normality of variables. To use these tests, the log of the data was tested for normality. If data has a multivariate normal distribution then each of the variables has a univariate normal distribution, but the opposite does not have to be true (MVN, 2015). For this reason, it is not enough to perform a

simple distribution check on each individual variable because we are interested in the joint distribution. Although this is the case, individual distribution checks were performed first to check whether each individual does follow a lognormal distribution.

A test of skewness was done for the individual distribution check which proved the distributions of individual variables to possess a lognormal distribution. After this simple test of skewness was complete, an in depth analysis was conducted to assess multivariate normality using a series of three statistical tests: Mardia's, Henze-Zirkler's, and Royston's tests.

Mardia's test is a popular method for determining multivariate skewness and kurtosis. The skewness ($\hat{\gamma}_{1,p}$) and kurtosis ($\hat{\gamma}_{2,p}$) are measured by (2-6) and (2-7) where $m_{ij} = (x_i - \bar{x})'S^{-1}(x_j - \bar{x})$, the squared Mahalanobis distance, and p is the number of variables.

$$\hat{\gamma}_{1,p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}^3 \tag{2-6}$$

$$\hat{\gamma}_{2,p} = \frac{1}{n}\sum_{i=1}^{n} m_{ii}^2 \tag{2-7}$$

The Henze-Zirkler test is based on the measured the distance between two distribution functions. The test statistic obtained is lognormally distributed if the data is normal. For the estimation of the p-value, the mean and variance are log-normalized. (MVN, 2015) The Henze-Zirkler test statistic is presented in (2-8) where p represents the number of variables. β is a function of the number of variables, and D_i and D_{ij} are the squared Mahalanobis distance of the ith observation to the centroid and the ith and jth observations, respectively.

$$HZ = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{-\frac{\beta^2}{2}D_{ij}} - 2(1+\beta^2)^{-\frac{p}{2}}\sum_{i=1}^{n} e^{-\frac{\beta^2}{2(1+\beta^2)}D_i} + n(1+2\beta^2)^{-\frac{p}{2}} \tag{2-8}$$

$$\beta = \frac{1}{\sqrt{2}}\left(\frac{n(2p+1)}{4}\right)^{\frac{1}{p+4}} \tag{2-9}$$

$$D_{ij} = \left(x_i - x_j\right)'^{S^{-1}}\left(x_i - x_j\right) \tag{2-10}$$

$$D_i = (x_i - \bar{x})'^{S^{-1}}(x_i - \bar{x}) = m_{ii} \tag{2-11}$$

For normally distributed multivariate data, the HZ test statistic is log normally distributed and therefore represented by two parameters: mean and variance. Thus, this test statistic for multivariate normality is calculated with (2-12).

$$z = \frac{\log(HZ) - \log(\mu)}{\log(\sigma)} \tag{2-12}$$

The last test used is Royston's test which uses the Shapiro-Wilk/Shapiro-Francia statistic, calculated using (2-13) where *e* is the equivalent degrees of freedom *(edf)* and *Φ(.)* is the cumulative distribution function for standard normal distribution with

$$H = \frac{e\sum_{j=1}^{p}\psi_j}{p} \sim X_e^2 \tag{2-13}$$

$$e = p/[1 + (p-1)\bar{c} \tag{2-14}$$

$$\psi_{ij} = \left\{ \Phi^{-1}\left[\frac{\Phi(-Z_j)}{2}\right]\right\}^2, j = 1,2,\dots,p \tag{2-15}$$

where c is a parameter that must be calculate from the correlation between variables, shown below.

$$c_{ij} = \begin{cases} g(r_{ij}, n) & if\ i \neq j \\ 1 & if\ i = j \end{cases} \tag{2-16}$$

$$g(r,n) = r^{\lambda}[1 - \frac{\mu}{\nu}(1-r)^{\mu} \tag{2-17}$$

The parameters μ, ν ,λ are estimated with μ=0.715 , λ=5 and (2-18) and (2-19) to calculate ν.

$$v(n) = 0.21364 + 0.015124x^2 - 0.0018034x^3 \tag{2-18}$$

$$x = \log(n) \tag{2-19}$$

Since there are three tests used here to confirm or deny multivariate normality, a Q-Q (quantile-quantile) plot is used to assist with any inconsistent results between the three tests. The Q-Q plot, is a commonly used graphical representation of the agreement between probability distributions. The Mahalanobis distance is a representation of the distance between a point and a

distribution. It is a multi-dimensional measure of how many standard deviations the point is from the mean of the distribution. One axis, the distance, represents theoretical (hypothesized) quantiles while the other axis represents the observed quantiles. The more of a fit there is between actual data and the assumed distribution, the better the points and the line will match. Figure 2.3 shows a Q-Q plot for generated multivariate normal data. As can be seen, the points closely match the line of best fit, indicating a multivariate normal distribution which we know to be true in this case.

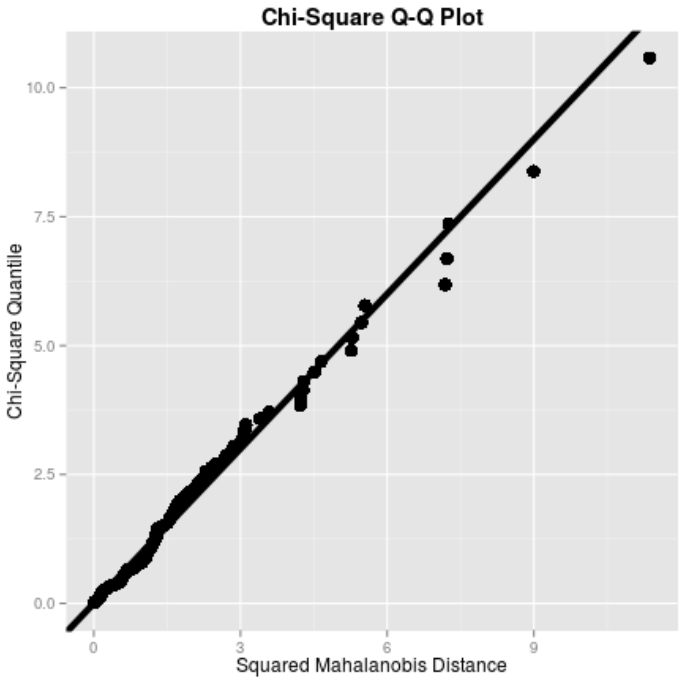


Figure 2.3: Normal Q-Q Plot

The Q-Q plots corresponding to each building for each hazard level are shown below. A large amount of deviation between the observed and expected data for the majority of cases considered, which suggests that the data is not normally distributed. For all three tests, the p values were significantly below 0.05, indicating no sign of multivariate normality. Two cases, the 4-story and 8-story building data for 50% in 50 years hazard, had a p-values of 0.05 and 0.08, respectively, for Royston's test. For the majority of cases, the variables do not appear to be jointly lognormal.

The following subsections display the detailed results from these multivariate normality tests for each case study building and hazard level considered.

2.5.1 4-story structure, 2% in 50 years hazard

For the 4-story building considering 2% in 50 years hazard, the Q-Q plot, presented in Figure 2.4, shows the deviation from multivariate normality of the data, especially at larger values of squared Mahalanobis distance. Detailed output results for each multivariate normality test (from R statistical analysis software) are provided in Appendix A. For all three tests, the p-values corresponding to the test statistics are significantly below 0.05, the chosen significance level, and therefore multivariate normality cannot be characteristic of the data.

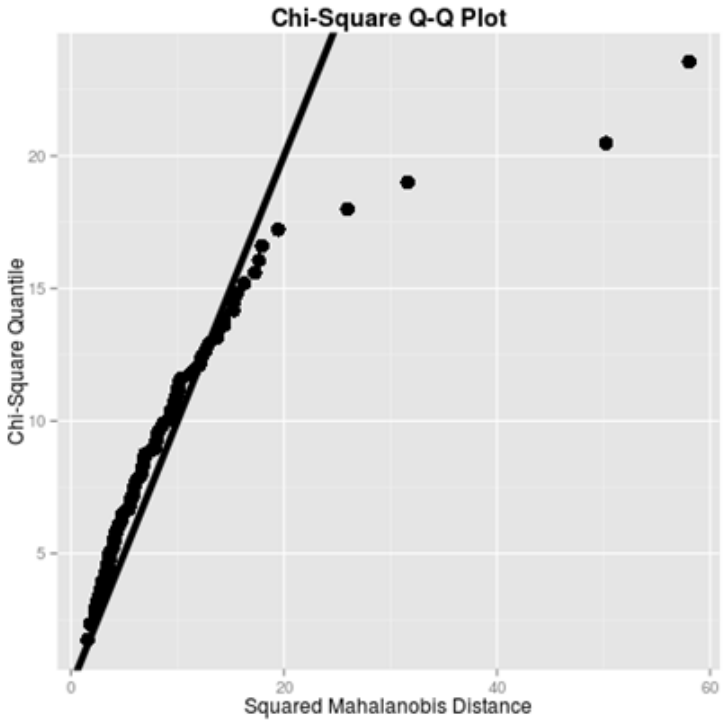


Figure 2.4: Q-Q plot for 4-story building considering 2% in 50 years hazard

2.5.2 4-story structure, 10% in 50 years hazard

The Q-Q plot for the 10% in 50 years hazard data (Figure 2.5) shows little variation between observed and expected data, especially compared to the 2% in 50 years data Q-Q plot. Similar to the 2% in 50 years data, all three multivariate normality tests for the 10% in 50 years hazard data is fail at 0.05 significance. All p-values, for all three tests, are far below the significance level, so even considering a stricter threshold (say alpha of 0.01), the data is still not considered to be multivariate normal.

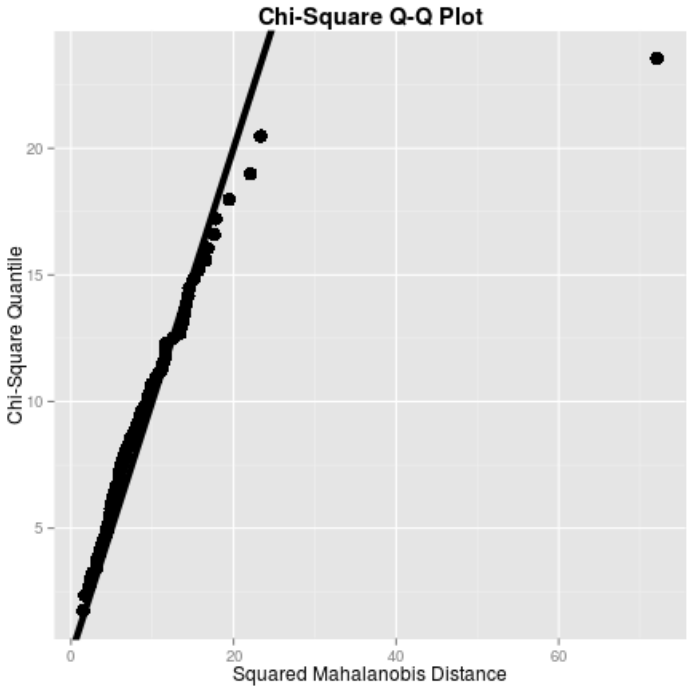


Figure 2.5: Q-Q plot for 4-story building considering 10% in 50 years hazard

2.5.2 4-story structure, 50% in 50 years hazard

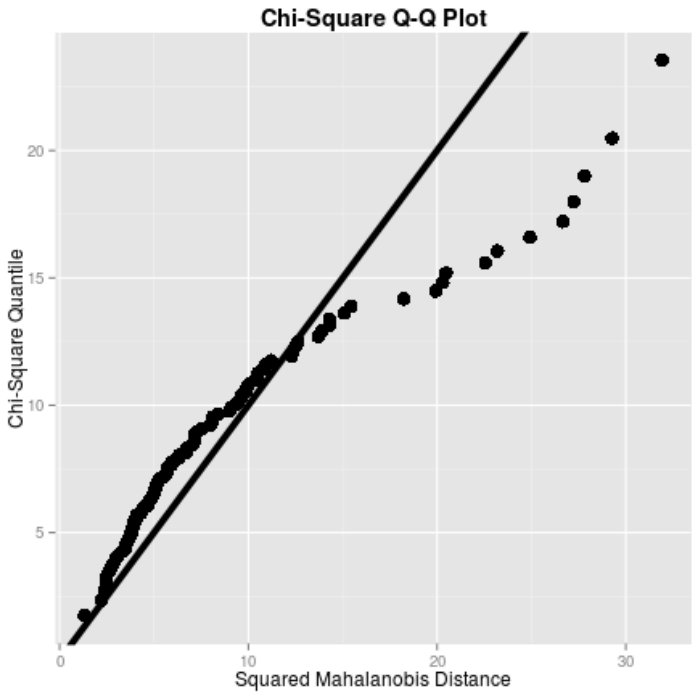


Figure 2.6: Q-Q plot for 4-story building considering 50% in 50 years hazard

This 4-story 50% in 50 years hazard data shows the largest deviation between observed and expected data on the Q-Q plot (Figure 2.6) compared to the other two considered hazard levels however, when results of Royston's multivariate normality test fail to reject that the data is multivariate normal. Royston's test statistic is 0.0598 (see Appendix A) which is very slightly above the chosen significance level of 0.05. For the other two tests, the data is shown to not be multivariate normal, with p-values significantly below alpha. Although Royston's test shows the data to be multivariate normal, considering the results from the other two tests and the shape of the Q-Q plot, the data is still deemed as not following multivariate normality. Therefore, it is argued that the 50% in 50 years hazard data for the 4-story building is not multivariate normal.

2.5.2 8-story structure, 2% in 50 years hazard

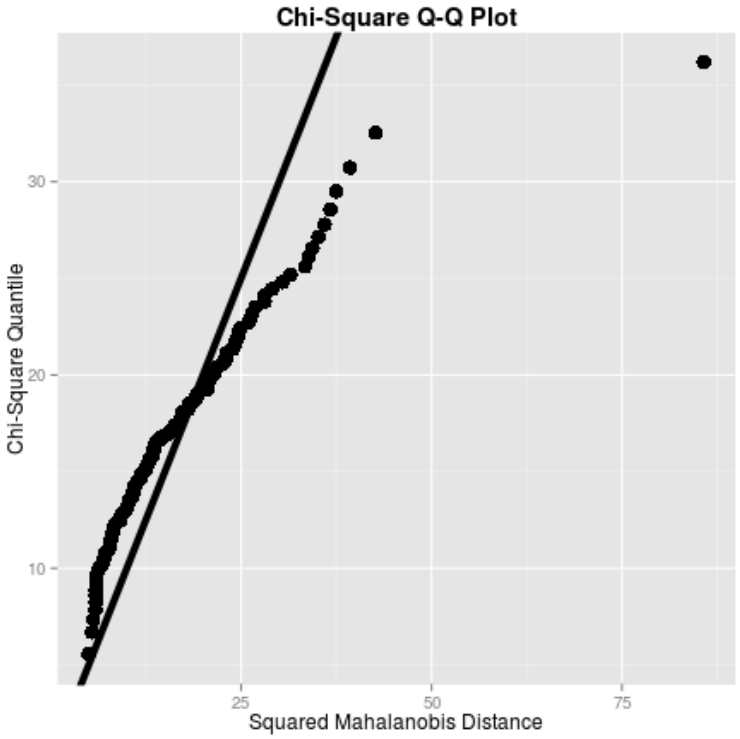


Figure 2.7: Q-Q plot for 8-story building considering 2% in 50 years hazard

The Q-Q plot for the EDP data of the 8-story structure considering 2% in 50 years hazard significantly deviates from the expected data in the Q-Q plot (Figure 2.7), which leads to a preliminary finding that the data is also, similar to the 4-story building at the 2% in 50 years hazard level, not multivariate normal. The three multivariate normality tests are used to further confirm or deny this result. For all three tests, the data is shown not to be multivariate normal. All of the test statistics correspond to p-values far below the chosen significance level of 0.05.

2.5.2 8-story structure, 10% in 50 years hazard

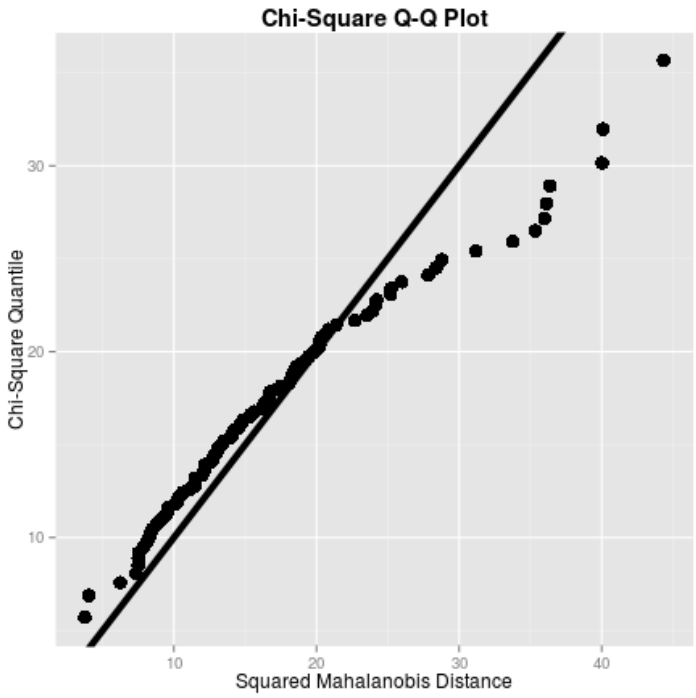


Figure 2.8: Q-Q plot for 8-story building considering 10% in 50 years hazard

The 8-story case study structure data for 10% in 50 years hazard also deviates from the expected values, which can be seen in Figure 2.8. For this plotted data, more deviation can be noticed along all values of distance (x-axis values) and all three multivariate normality tests fail to show that the data is multivariate normal.

2.5.2 8-story structure, 50% in 50 years hazard

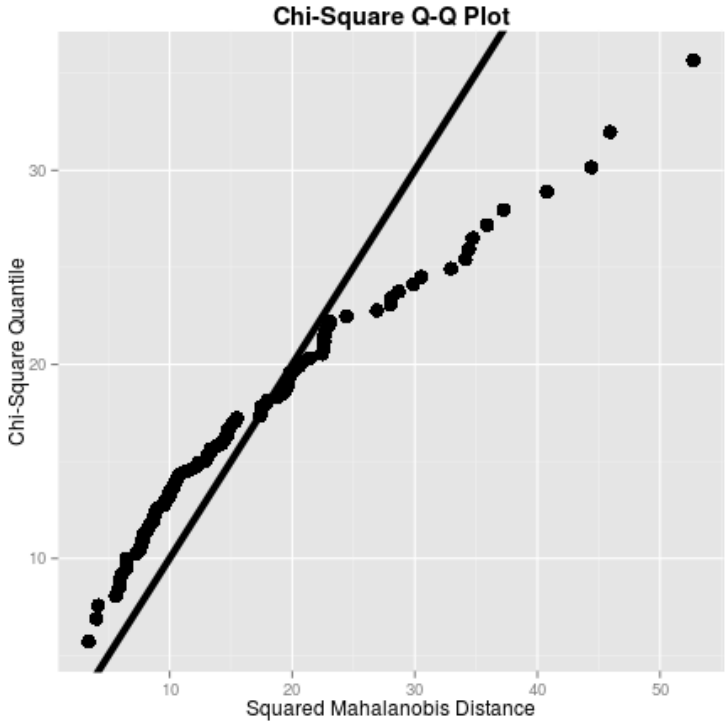


Figure 2.9: Quantile-quantile plot for 8-story building considering 50% in 50 years hazard

The Q-Q plot for the 8-story structure data considering 50% in 50 years hazard is shown in Figure 2.9. A significant amount of deviation can be seen throughout the plot, which is slightly different than the previous data, specifically for the 2-story building 2% and 10% in 50 years hazard data. As building story increases and probability of exceedance increases, the less of a fit there is between the observed and expected data on the Q-Q plot.

Similar to the results of the 4-story building when considering 50% in 50 years hazard, Royston's test concludes that the 8-story building 50% in 50 years hazard data is multivariate normal. The p-value for the 8-story building (0.08) is slightly higher than that of the 4-story building (0.06), but they are both very close to the chosen significance level of 0.05, which means

that they barely passed the threshold of significance. This test result is rejected when considering the other two tests and the chi-squared plot.

From the above results it is determined that at a 5% significance level, EDPs, for both structures and at each considered hazard level, do not follow a joint lognormal distribution. Firstly, judging by the Q-Q plots that were presented, a large amount of deviation was present between the observed and expected data, suggesting that the data is not multivariate normal. Secondly, for all three tests and for the majority of cases presented, the p-values of the test statistics were significantly below alpha (0.05) which rejects the claim that the data is multivariate normal. Only two cases presented resulted in a p-value greater than alpha which was for Royston's test statistics for both the 4- and 8- story building when considering the 50% in 50 years hazard data. The p-value was on the border and barely passed at 0.05 significance for both cases. This result can be disregarded when looking at the results from the Q-Q plot, Mardia's test and the Henze-Zirkler test which all showed that the data is not multivariate normal. The purpose of conducting three tests and confirming these results with the Q-Q plot is so that in times when one test suggests something different from another, the majority of test results can be used to draw the most accurate conclusion. Therefore, based on these three multivariate normal tests and the Q-Q plots, all of the considered data is argued to follow a multivariate distribution other than a lognormal distribution.

2.6 EDP SAMPLES GENERATED USING COPULAS AND JOINT LOGNORMAL DISTRIBUTION

In order to investigate the accuracy of a EDPs generated using different numbers of initial data points, random samples with 5, 11, 15, 25, 50 and 100 initial data points, sampled from the population EDPs, are used to obtain 100 realizations of EDPs per analysis. For example, 5 rows of EDPs are randomly sampled from the population of 100 EDPs and used to generate 100 copula EDPs and 100 joint lognormal EDPs, which is done 10 times per initial sample size (5, 11, 15 ,25, 100), per case study SMRF (2-, 4-, 8-, 12- story), per hazard level (10% and 50% in 50 years). A Kolmogorov-Smirnov (KS) goodness of fit test is used to measure the equality of the sample distributions with the populations they are drawn from. A KS test is chosen as it is a nonparametric test that does not require the sample or population to be from a standard distribution, unlike other commonly use hypothesis tests such as Student's t-test. The KS statistic measures the greatest distance between two empirical distribution functions and is calculated by Equation (2-20) where D_{mn} is the KS test statistic, $\underset{x}{sup}$ denotes the supremum, $F_m(EDP_i)$ is the cumulative distribution function of each population EDP_i, and $G_n(EDP_i)$ is the cumulative distribution function for each sample EDP_i.

$$D_{mn} = \underset{EDP_i}{sup} |F_m(EDP_i) - G_n(EDP_i)| \tag{2-20}$$

The null hypothesis, H_0, states that the sample data follows the same distribution as the population while the alternative hypothesis, H_a, states that the sample data does not follow the same distribution as the population. 10 random samples from the population of EDPs corresponding to each sample size are taken for a total of 60 random samples (6 different initial sample sizes used) per hazard level per case study SMRF. For each random sample that is drawn from the population EDPs, this same random sample is used to generate 100 EDPs using 1) Gaussian copulas and 2) joint lognormal distribution. Tables 3 through 6 present the results of the

KS tests, where each number in the table corresponds to the percentage of failures of the 10 random samples for each considered sample size. The darker the grey shade, the larger the number of failures for that particular sample size and EDP. As the number of stories increases, the minimum required sample size to see a majority of sample distributions that match the population distributions increases. For the 2-story SMRF, the highest percentage of observed failures for any case is 30%, and there are several cases where all samples tested match the population distribution. For the 8-story building, there are several cases that fail 60% of the time with smaller sample sizes, especially when using a joint lognormal distribution to represent EDP dependence even when a full rank covariance matrix is used. There are several other cases where, even with a full rank covariance matrix, joint lognormal EDPs exhibit 40% or more failure. For example, using a sample size of 15 for the 4-story SMRF, which is greater than the number of EDPs (9 EDPs for the 4-story SMRF) by 6, still yields 40% failure is several cases at both hazard levels when a joint lognormal distribution is used. Even with increasing the sample size to 25 in the case of 10% hazard yields in several instances of 40% failure to match the population distribution. In contrast, copula EDPs only have one instance of failure at 40% for each hazard level for the 4-story SMRF considering a sample size of 15, with all other EDPs showing 30% or less failure and for a sample size of 25, the majority of cases are around 10-20% failure. From Tables 3 through 6 it can be seen that for the majority of cases, EDPs generated using Gaussian copulas are better able to match the population EDPs compared with EDPs generated using a joint lognormal distribution.

2-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	10	10	20	10	0	10
PFA1	20	10	10	10	0	10
PFA2	30	20	10	10	20	0
IDR1	10	20	10	10	10	0
IDR2	10	10	20	10	0	0

2-story SMRF Copula Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	10	20	0	0	10	0
PFA1	10	20	10	10	0	0
PFA2	20	20	20	0	0	0
IDR1	20	10	0	0	10	10
IDR2	10	10	0	20	0	0

2-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	20	10	10	10	10	0
PFA1	20	10	0	20	10	10
PFA2	30	20	20	20	20	10
IDR1	10	20	20	10	10	0
IDR2	10	30	30	20	10	0

2-story SMRF Copula Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	10	10	0	0	0	0
PFA1	20	10	0	0	0	10
PFA2	20	10	10	20	0	0
IDR1	20	20	0	20	0	0
IDR2	10	10	0	10	0	0

Figure 2.10: Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 2-story SMRF and 10% and 50% in 50 years hazard with the grey scale indicating levels of failure

4-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	20	20	40	40	20	10
PFA1	50	40	30	30	10	10
PFA2	50	30	50	50	40	30
PFA3	30	20	20	20	30	30
PFA4	30	30	30	40	10	30
IDR1	40	30	30	10	10	30
IDR2	50	30	30	30	20	30
IDR3	20	20	20	40	20	30
IDR4	40	30	30	40	20	30

4-story SMRF Copula Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	50	40	30	10	0	0
PFA1	50	30	10	30	0	0
PFA2	10	20	30	10	10	10
PFA3	20	40	20	30	0	0
PFA4	20	20	30	40	20	0
IDR1	50	30	20	10	10	10
IDR2	40	40	30	10	20	10
IDR3	50	20	40	20	30	30
IDR4	40	40	20	20	10	30

4-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	40	40	40	30	30	0
PFA1	50	30	40	30	30	20
PFA2	50	40	30	30	10	20
PFA3	40	30	30	30	20	30
PFA4	30	30	30	20	10	20
IDR1	30	30	40	20	10	30
IDR2	40	20	40	20	20	10
IDR3	20	30	40	20	20	20
IDR4	30	20	20	30	30	20

4-story SMRF Copula Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	50	20	10	40	10	10
PFA1	20	20	10	20	10	0
PFA2	50	30	20	20	10	0
PFA3	40	50	30	20	20	10
PFA4	30	40	20	10	0	0
IDR1	60	20	30	20	0	30
IDR2	60	20	30	10	10	20
IDR3	20	20	30	10	20	10
IDR4	30	30	40	10	20	20

Figure 2.11: Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 4-story SMRF and 10% and 50% in 50 years hazard with the grey scale indicating levels of failure

8-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	70	40	50	30	30	20
PFA1	70	50	50	30	30	20
PFA2	60	40	40	20	20	20
PFA3	30	40	30	30	40	20
PFA4	30	50	20	30	30	10
PFA5	30	40	20	10	30	20
PFA6	50	30	40	20	30	10
PFA7	50	40	50	20	40	30
PFA8	50	40	50	30	30	30
IDR1	60	40	50	10	20	30
IDR2	60	50	50	40	30	30
IDR3	60	50	50	40	30	30
IDR4	40	50	40	30	30	30
IDR5	40	50	50	40	30	10
IDR6	60	70	60	50	40	40
IDR7	60	40	40	30	30	30
IDR8	60	30	50	50	40	30

8-story SMRF Copula samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	50	40	30	20	10	0
PFA1	50	50	30	20	10	0
PFA2	50	40	30	10	10	0
PFA3	60	20	50	20	10	10
PFA4	40	20	30	30	10	0
PFA5	40	20	30	20	20	20
PFA6	40	30	40	20	20	10
PFA7	40	40	30	20	30	10
PFA8	40	20	30	20	10	20
IDR1	40	30	20	30	20	10
IDR2	60	40	40	20	20	20
IDR3	40	40	40	10	10	30
IDR4	40	30	30	10	10	30
IDR5	40	40	30	10	20	20
IDR6	40	60	20	20	30	10
IDR7	40	40	20	30	20	20
IDR8	50	40	20	20	10	20

8-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	30	40	20	20	10	20
PFA1	30	40	40	20	10	20
PFA2	40	30	30	20	20	30
PFA3	30	30	40	30	40	10
PFA4	50	40	40	10	30	10
PFA5	60	50	50	40	20	0
PFA6	60	40	40	40	20	20
PFA7	50	50	40	20	10	30
PFA8	60	30	30	30	40	30
IDR1	60	30	30	40	20	20
IDR2	60	30	40	30	20	20
IDR3	50	30	40	30	20	10
IDR4	50	40	50	30	10	20
IDR5	60	40	40	40	20	20
IDR6	60	40	30	30	20	10
IDR7	40	40	40	20	30	10
IDR8	50	50	40	20	30	20

8-story SMRF Copula samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	60	30	30	10	10	10
PFA1	50	50	30	20	10	10
PFA2	50	40	30	20	10	10
PFA3	60	40	30	40	10	30
PFA4	60	40	20	20	20	30
PFA5	40	40	40	30	30	20
PFA6	40	30	40	30	20	30
PFA7	70	40	40	30	20	0
PFA8	60	40	30	40	10	20
IDR1	40	50	40	30	20	20
IDR2	40	40	40	20	30	10
IDR3	40	60	30	20	30	10
IDR4	50	30	20	10	50	10
IDR5	60	20	30	30	20	10
IDR6	60	20	20	30	20	0
IDR7	50	40	20	10	10	10
IDR8	60	40	20	20	20	10

Figure 2.12: Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 8-story SMRF and 10% and 50% in 50 years hazard with the grey scale indicating levels of failure

12-story SMRF Joint Lognormal Samples 10% in 50 years

Sample Size	5	11	15	25	50	100
PGA	40	60	50	60	30	30
PFA1	40	60	40	60	40	30
PFA2	50	60	40	60	40	30
PFA3	60	50	50	50	50	20
PFA4	60	50	40	60	20	20
PFA5	60	60	50	50	20	10
PFA6	40	60	50	30	10	20
PFA7	60	70	60	50	20	20
PFA8	70	50	40	50	30	20
PFA9	60	50	30	60	30	30
PFA10	70	40	40	30	10	30
PFA11	70	40	30	30	30	10
PFA12	70	40	30	40	20	20
IDR1	50	50	40	40	50	10
IDR2	60	60	30	30	20	20
IDR3	60	40	50	30	20	20
IDR4	40	40	50	30	30	20
IDR5	40	40	40	50	40	30
IDR6	50	30	40	30	40	20
IDR7	40	40	50	40	10	10
IDR8	50	40	30	40	10	10
IDR9	60	50	30	30	30	20
IDR10	70	60	40	40	20	20
IDR11	70	50	30	40	20	10
IDR12	70	60	30	50	20	30

12-story SMRF Copula samples 10% in 10 years

Sample Size	5	11	15	25	50	100
PGA	60	50	20	50	40	10
PFA1	70	50	40	20	30	10
PFA2	60	50	50	10	30	0
PFA3	50	50	40	50	20	30
PFA4	60	60	30	40	40	20
PFA5	50	50	40	30	10	20
PFA6	30	50	40	30	20	10
PFA7	40	40	40	30	10	20
PFA8	40	40	30	30	30	20
PFA9	30	40	20	30	30	0
PFA10	50	50	40	40	30	10
PFA11	40	40	30	20	30	30
PFA12	40	50	50	40	20	10
IDR1	40	50	40	50	40	20
IDR2	50	50	60	20	40	30
IDR3	60	60	50	20	30	30
IDR4	30	50	60	10	30	10
IDR5	40	40	40	20	0	0
IDR6	50	30	40	50	10	10
IDR7	60	50	20	30	20	20
IDR8	50	50	60	10	10	20
IDR9	40	40	40	50	30	0
IDR10	50	40	40	40	30	10
IDR11	60	50	50	20	20	10
IDR12	60	40	40	50	20	30

12-story SMRF Joint Lognormal Samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	80	50	20	50	30	20
PFA1	60	40	50	30	40	10
PFA2	60	50	50	40	40	20
PFA3	30	60	60	50	30	20
PFA4	40	60	40	40	20	20
PFA5	30	70	50	20	30	30
PFA6	50	50	60	60	30	20
PFA7	50	60	50	50	0	10
PFA8	50	70	30	40	30	10
PFA9	40	70	40	50	20	20
PFA10	70	70	20	50	10	10
PFA11	60	40	60	30	40	20
PFA12	80	50	40	40	20	30
IDR1	60	60	40	40	20	30
IDR2	50	50	30	30	30	30
IDR3	60	50	50	30	40	20
IDR4	40	60	40	40	40	30
IDR5	40	40	30	50	20	10
IDR6	50	30	40	20	50	30
IDR7	50	60	50	40	40	20
IDR8	50	60	60	40	20	10
IDR9	60	70	30	30	10	10
IDR10	60	50	30	30	20	30
IDR11	50	40	40	50	20	20
IDR12	60	50	50	60	30	20

12-story SMRF Copula samples 50% in 50 years

Sample Size	5	11	15	25	50	100
PGA	60	70	40	30	10	0
PFA1	60	40	50	50	30	30
PFA2	70	60	60	20	40	20
PFA3	60	60	50	20	20	20
PFA4	50	50	50	40	20	10
PFA5	60	50	60	30	0	30
PFA6	70	70	40	10	10	20
PFA7	40	50	30	30	40	30
PFA8	80	40	30	20	20	0
PFA9	50	30	30	30	30	10
PFA10	60	50	40	10	0	20
PFA11	70	60	50	20	30	20
PFA12	40	60	40	30	20	10
IDR1	60	70	40	30	20	0
IDR2	50	50	60	40	10	10
IDR3	70	70	40	60	30	10
IDR4	50	40	40	30	30	10
IDR5	60	50	40	30	40	20
IDR6	50	60	40	40	20	30
IDR7	40	60	50	30	20	30
IDR8	50	50	30	40	10	30
IDR9	70	50	60	20	20	30
IDR10	50	70	20	20	20	20
IDR11	70	60	50	10	10	10
IDR12	70	40	20	40	10	20

Figure 2.13: Percentage of failures for KS tests with multiple random samples of EDPs that are used to generate copula and joint lognormal EDPs for 12-story SMRF and 10% and 50% in 50 years hazard with the grey scale indicating levels of failure

2.9 RESULTS

This study presents an approach, using Gaussian copulas, to model the dependence of engineering demand parameters used in performance assessment of buildings subject to earthquake hazards. Current provisions outlined by FEMAP-58 (2015) provide a methodology that assumes EDPs follow a joint lognormal distribution. Using Gaussian copulas allows for the generation of a suite of EDPs from a smaller number of initial realizations without the need for potentially inaccurate assumptions regarding EDP dependence. Peak floor acceleration and maximum interstory drift ratios for four special steel moment resisting frame buildings are obtained from a large suite of selected and scaled ground motion records. These values of demand are then used to generate EDPs that exhibit the dependence structure of Gaussian copulas and also, that follow a joint lognormal distribution, which follow the current methodology outlined by FEMAP-58 (2015). Results show that EDPs generated using Gaussian copulas are better able to match the population EDPs from which they were sampled, especially with smaller initial input realizations, compared with EDPs generated that follow a joint lognormal distribution.

2.10 REFERENCES

AISC, A. (2005). 360-Specification for Structural Steel Buildings; American Institute of Steel Construction.

ASCE (2006) Seismic Rehabilitation of Existing Buildings, ASCE/SEI 41-06, American Society of Civil Engineers, Reston, Virginia

ASCE (2010) Minimum Design Loads for Buildings and Other Structures, ASCE/SEI 7-10, American Society of Civil Engineers, Reston , Virginia

Aslani, H. and Miranda, E. (2005), *Probabilistic Earthquake Loss Estimation and Loss Disaggregation in Buildings*. Diss. Stanford University.

ATC - Applied Technology Council, FEMA P-58 [2015b] *Next-generation Seismic Performance Assessment for Buildings, Volume 2 – Implementation Guide*, Federal Emergency Management Agency, Washington, D.C., 2015.

Baker, J. W. (2010). Conditional mean spectrum: Tool for ground-motion selection. *Journal of Structural Engineering*, *137*(3), 322-331.

Devroye, L. (1989). A note on Linnik's distribution. *Statistics & probability letters*, *9*(4), 305-306.

FEMA (2015) Next-Generation Methodology for Seismic Performance Assessment of Buildings, prepared by the Applied Technology Council for the Federal Emergency Management Agency, Report No. FEMA P-58, Washington, D.C.

Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, *12*(4), 347-368.

Goda, K. (2010). Statistical modeling of joint probability distribution using copula: application to peak and permanent displacement seismic demands. *Structural Safety*, *32*(2), 112-123.

Goda, K., & Tesfamariam, S. (2015). Multi-variate seismic demand modelling using copulas: Application to non-ductile reinforced concrete frame in Victoria, Canada. *Structural Safety*, *56*, 39-51.

Habiboellah, F. (2007). Copulas, Modeling dependencies in Financial Risk Management. *Purmerend, Dec*.

Jaworski, P. (2009). On copulas and their diagonals. *Information Sciences*, *179*(17), 2863-2871.

Jayaram, N., Lin, T., & Baker, J. W. (2011). A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. *Earthquake Spectra*, *27*(3), 797-815.Huang et al, 2011

Jolliffe, I. (2011). *Principal component analysis* (pp. 1094-1096). Springer Berlin Heidelberg.

Miranda, E., Aslani, H., & Taghavi, S. (2003). Assessment of seismic performance in terms of economic losses. In *Proceedings, International Workshop on Performance-Based Seismic Design: Concepts and Implementation* (Vol. 28, pp. 149-160). Berkeley: Pacific Earthquake Engineering Research (PEER) Center, University of California.

Moehle, J., & Deierlein, G. G. (2004, August). A framework methodology for performance-based earthquake engineering. In *13th world conference on earthquake engineering* (Vol. 679).

Nelsen, R. B. (1994), "Introduction." *An Introduction to Copulas*. Springer New York, 1-4.

Olshansky, R. B., & Wu, Y. (2001). Earthquake risk analysis for Los Angeles County under present and planned land uses. *Environment and Planning B: Planning and Design*, *28*(3), 419-432.

Sasani, M., & Kiureghian, A. D. (2001). Seismic fragility of RC structural walls: displacement approach. *Journal of Structural Engineering*, *127*(2), 219-228.

Shome, N., & Cornell, C. A. (1999). Probabilistic seismic demand analysis of nonlinear structures, Report No. RMS-35. *Research Management System (RMS) Program, Department of Civil Engineering, Stanford University, Stanford, CA*.

Shinozuka, M., Feng, M. Q., Lee, J., & Naganuma, T. (2000). Statistical analysis of fragility curves. *Journal of engineering mechanics*, *126*(12), 1224-1231.

Song, J., & Ellingwood, B. R. (1999). Seismic reliability of special moment steel frames with welded connections: II. *Journal of Structural Engineering*, *125*(4), 372-384.

Team, R. C. (2013). R: A language and environment for statistical computing.

Yang, T. Y., Moehle, J., Stojadinovic, B., & Der Kiureghian, A. (2006, April). An application of PEER performance-based earthquake engineering methodology. In *Proceedings*.

Yang TY, Moehle JP, Stojadinovic B, Der Kiureghian A (2009) Seismic performance evaluation of facilities: methodology and implementation. J Struct Eng-ASCE 135(10):1146–1154

CHAPTER 3: INCREASED ACCURACY OF SEISMIC PERFORMANCE ASSESSMENT THROUGH THE UTILIZATION OF COPULAS TO GENERATE ENGINEERING DEMAND PARAMETERS

3.1 INTRODUCTION

The performance of buildings subject to earthquakes can be measured in terms of several different measures which are intended to be meaningful to key stakeholder and decision makers. One of the most commonly used performance measures is economic loss. This chapter of the dissertation extends the work conducted in Chapter 2 to assess how different methodologies to generate EDPs correlate with differences at the loss level.

Differences in realizations of peak floor acceleration (PFA) and maximum interstory drift ratio (MaxIDR) using copulas and the joint lognormal distribution assumption to represent dependence are investigated. These two EDPs are chosen because of their respective influence on acceleration and drift, therefore encompassing a demand vector capable of presenting a holistic overview of response. Furthermore, a Kolmogorov-Smirnov test is conducted to determine the optimal number of initial ground motions required for accurate assessments of response when EDPs are generated using copulas and also assuming a joint lognormal distribution. Copulas have been used in a wide variety of applications, especially in the field of financial risk management to measure risk across a diverse range of areas in finance (Habiboellah, 2007), as they provide a convenient way to model the dependence between multi-variate data (Goda and Tesfamariam, 2015; Genest and Favre, 2007). Both of these methods for generating demand realizations are generating seismic loss are applied to a 2-, 4-, 8- and 12-story special steel moment resisting frame buildings located in Los Angeles, California, a high seismicity region with expected earthquake risk of hundreds of millions of dollars per year (Olshanksky and Yueming, 2001). This study contributes to the field of performance based earthquake engineering by providing an alternative statistical method for the

generation of PFA and IDR in performance assessment methodology that is shown to be more accurate in predicting economic losses incurred for buildings subject to earthquakes.

3.2 ESTIMATION OF ECONOMIC LOSS VIA FEMA PERFORMANCE ASSESSMENT CALCULATION TOOL

The Performance Assessment Calculation Tool (PACT), developed by the Applied Technology Council (ATC), is an computer-based calculation tool which includes a repository of fragility and consequence data to perform probabilistic calculations and accumulation of losses according to the methodology described in FEM P-58-1 (2015). A PACT model is developed for each case study SMRF, and several tests are run to generate the economic loss associated with the input demand matrices. Collapse cases are not considered for these analyses as the aim of this study is to assess the effect of different methods for generating EDPs on estimated of loss, not on whether or not the buildings collapse. Moreover, the probability of collapse of the buildings used in this study is insignificant at the studied hazard levels and chances of hitting a collapse damage state is highly unlikely. Therefore, realizations indicating a collapse limit state are limited to a couple, if any, which would only minutely affect one tail of the loss distribution. For each case study SMRF, the *population* EDPs for the 10% and 50% hazard levels is input into PACT to generate losses (herein referred to as *population* loss). Along with *population* loss, the economic loss associated with 100 joint lognormal (herein referred to as *lognormal* loss) and 100 *copula* EDPs (herein referred to as *copula* loss) is also generated. Based on the results from the previous section, for each building, enough initial realizations are used to create a full rank covariance matrix. Therefore, for the 2- and 4-story buildings, 11 initial *population* EDPs are used to generate 100 *copula* and *lognormal* EDPs and for the 8- and 12-story buildings, 25 and 50 initial *population* EDPs are used to generate 100 *copula* and *lognormal* EDPs, respectively. All of the structural and

nonstructural components in the case study SMRFs are defined by fragility and loss functions based on the normative values provided in Appendix F of FEMA P-58-1 (2015) for the 'Research' occupancy category. All values of monetary loss resulting from PACT analyses were normalized by average maximum loss of damageable components, which was calculated per building by pushing the buildings to maximum capacity, forcing them to the largest possible damage state. The following are the monetary values of average maximum loss, in millions of dollars, for the 2-, 4-, 8- and 12- story buildings, respectively: 7.46, 14.92, 29.84, 44.76.

The results of these analyses are shown in Figures 1-4, where the black asterisk curves represent the CDFs of *population* loss and the red dotted curves represent the CDFs of the *copula* and *lognormal* EDPs, respectively.

For the 2-story SMRF, both *copula* and *lognormal* EDPs perform well in estimating loss considering 10% in 50 years hazard, but for the 50% in 50 years hazard level, *copula* EDPs are capable of replicating losses generated by the *population* EDPs. While, in general, both methods are acceptable in matching loss generated from *population* EDPs, for the majority of cases, *copula* EDPs are superior to *lognormal* EDPs with the exception of the 4-story building and 8-story at 10% in 50 years hazard. In the latter, *lognormal* EDPs generate loss values that are better at matching the *population* loss. It is worth noting that as story height increases, a smaller difference in loss is noticed between the two ground motion hazard levels, which is especially noticeable for the 12-story building. The demand for the higher rise systems was relatively small which caused most of the damage to fall within the same damage state for both hazard levels, which then corresponds to similar values of loss, regalrdess of ground motion intensity. To quantify the difference between distribution of losses obtained from *copula* and *lognormal* EDPs, with *population* loss, respectively, Kullback-Leibler (KL) divergence is used. KL divergence measures

the distance between one probability distribution and another reference distribution, is calculated. A KL divergence of 0 corresponds to two identical distributions. Let P and Q be discrete probability distributions and D_{KL} be the KL divergence between P and Q, represented by Equation (8).

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log\left(\frac{Q(x)}{P(x)}\right) \quad (8)$$

The KL divergence of each *copula* and *lognormal* loss CDF and *population* loss CDF are calculated. For each hazard level and SMRF, the average KL divergence is calculated and presented in Table 1. For example, the average KL divergence of all 10 generated *copula* loss samples with *population* loss for the 2-story building is 0.042. In all but one case (12-story 10% in 50 years hazard) *copula* loss has a smaller KL divergence compared with *lognormal* loss. These findings are in line with previous results which showed that *copula* EDPs are able to better match *population* EDPs at smaller sample sizes in most cases, especially when a full rank covariance matrix is achieved.

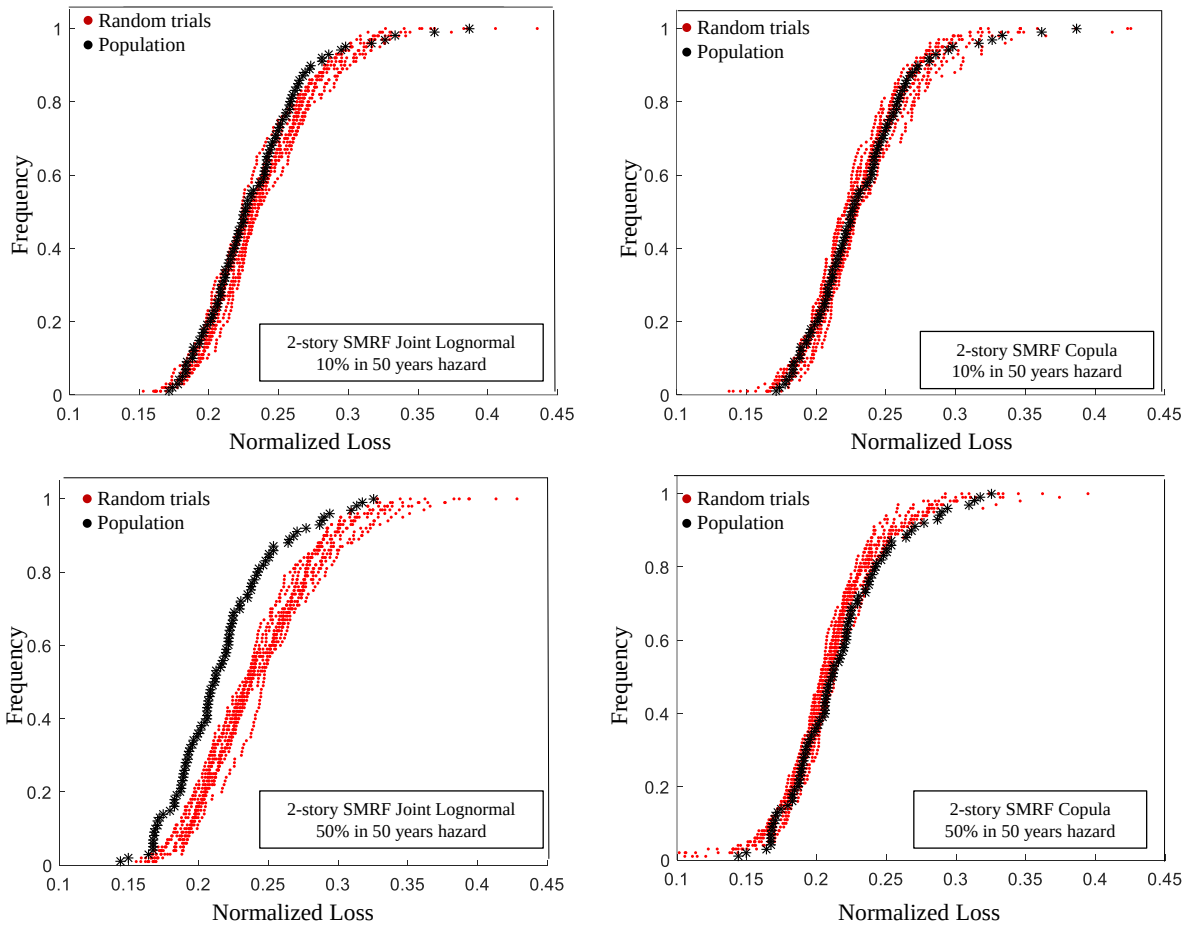


Figure 3.1: Cumulative distribution functions of loss normalized to average maximum loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 2-story SMRF considering 10% and 50% in 50 years hazard

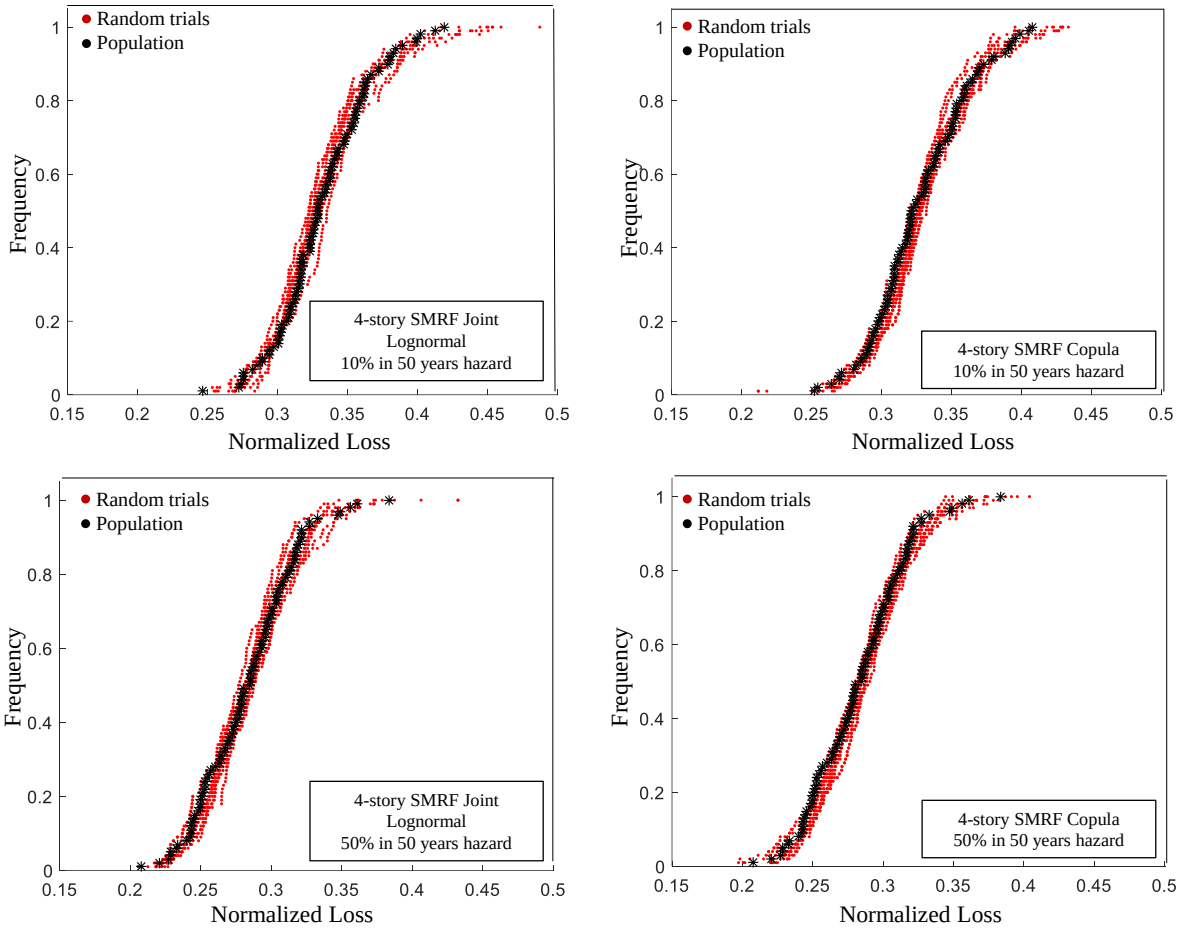


Figure 3.2: Cumulative distribution functions of loss normalized to average maximum loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 4-story SMRF considering 10% and 50% in 50 years hazard

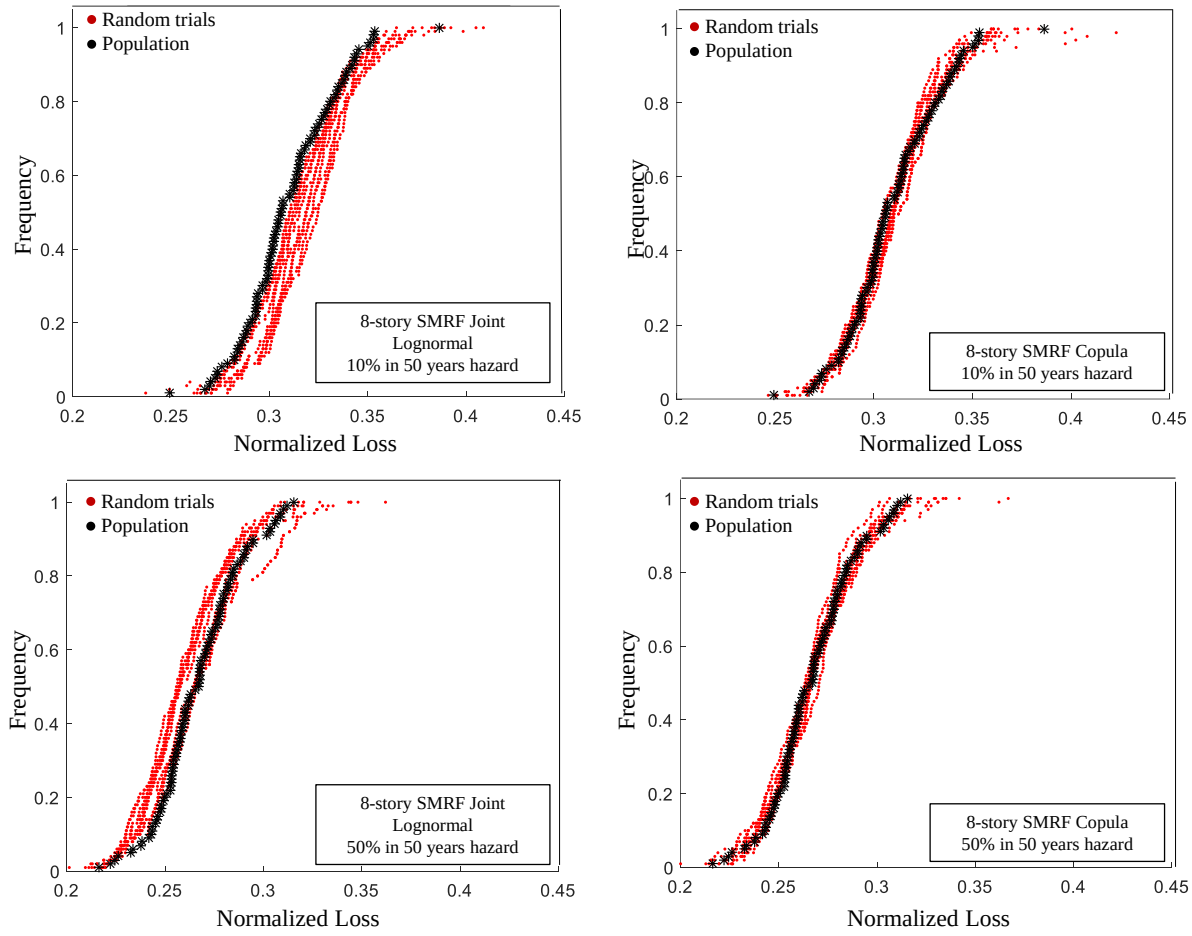


Figure 3.3: Cumulative distribution functions of loss normalized to average maximum loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 8-story SMRF considering 10% and 50% in 50 years hazard

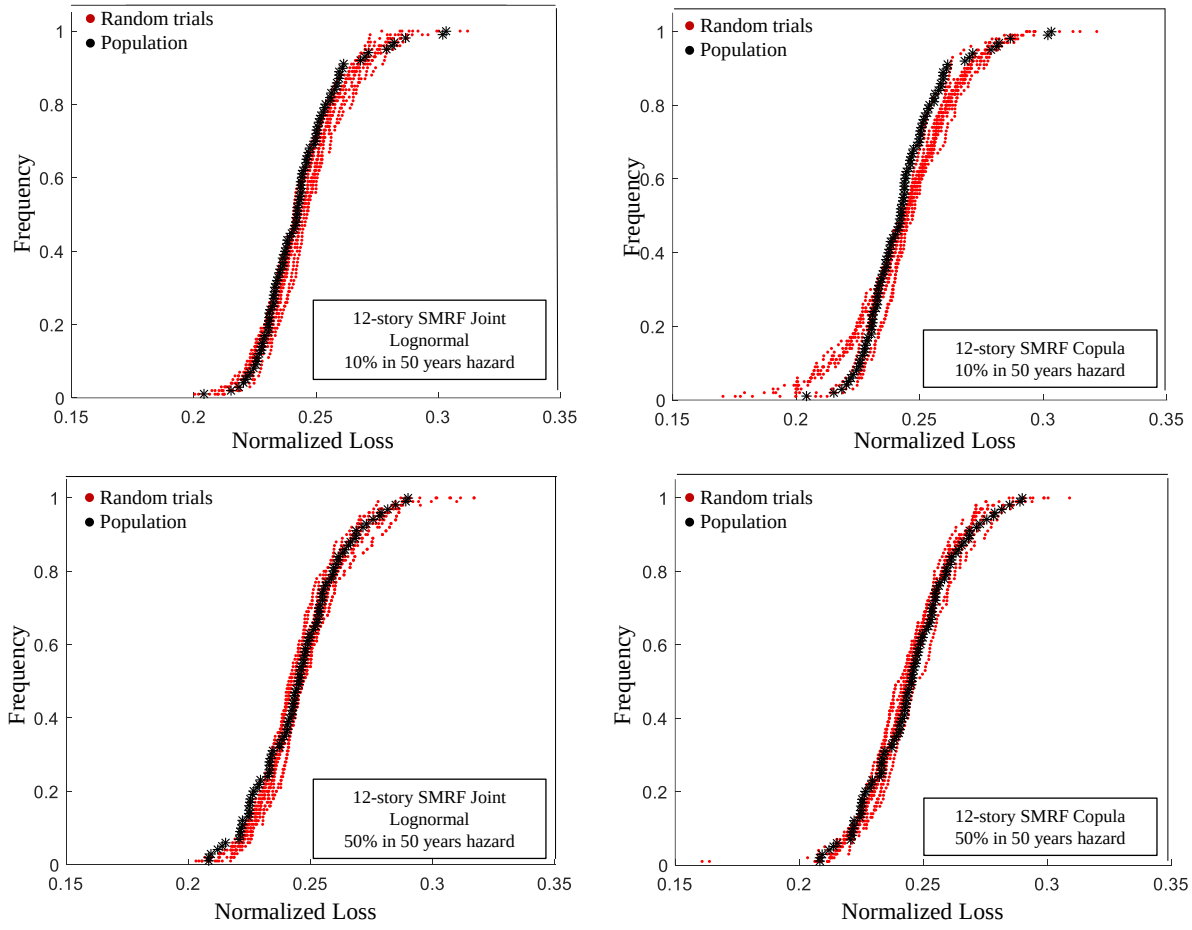


Figure 3.4: Cumulative distribution functions of loss normalized to average maximum loss generated using PACT for population EDPs and EDPs generated using copula and joint lognormal for the 12-story SMRF considering 10% and 50% in 50 years hazard

Table 3.1: Average KL Divergence values between population and simulation CDFs for 10% and 50% probability of exceedance in 50 years

	Lognormal (10% in 50 years)	Copula (10% in 50 years)	Lognormal (50% in 50 years)	Copula (50% in 50 years)
2-story	0.0196	0.0186	0.0617	0.0237
4-story	0.025	0.0161	0.0145	0.0132
8-story	0.0475	0.0301	0.0142	0.0125
12-story	0.0118	0.0123	0.0183	0.0116

3.3 CONCLUSION

This study presents an approach, using Gaussian copulas, to model the dependence of engineering demand parameters used in performance assessment of buildings subject to earthquake hazards. Current provisions outlined by FEMAP-58 (2015) provide a methodology that assumes EDPs follow a joint lognormal distribution. Using Gaussian copulas allows for the generation of a suite of EDPs without the need for potentially inaccurate assumptions regarding EDP dependence. Peak floor acceleration and maximum interstory drift ratios for four special steel moment resisting frame buildings are obtained from a large suite of selected and scaled ground motion records. These values of demand are then used to generate EDPs that exhibit the dependence structure of Gaussian copulas and also, that exhibit a joint lognormal distribution, which follows the current methodology outlined by FEMAP-58 (2015). Results show that for lower rise buildings (i.e. 2- and 4-story systems) approximately 30-40% accuracy (accuracy measured in terms of statistical similaties between simulated and observed demand sets) in simulated demand sets can be achieved though assuming a joint lognormal distribution even when the initial sample size is small. However, for the lower rise buildings, using Gaussian copulas yields several instances where simulated demand is 90-100% accurate, but a larger number of initial observations must be used to achieve this accuracy. In constrast, even with more initial observations, EDPs generated by assuming a joint lognormal distribution still do not achieve the same level of accuracy as EDPs generated using Gaussian copulas. In higher rise buildings (i.e. 8- and 12-story systems), when

generating EDPs using sample sizes that are significantly smaller than the number of variables being generated results are innacruate for EDPs generated using both Gaussian copulas and a joint lognormal distribution. However, as sample size increases, there is a clear trend of significantly increasing accuracy in EDPs generated using Gaussian copulas, but this trend is not as strong for EDPs generated using a joint lognormal distribution. In summary, using copulas along with enough initial obsevations to achieve a full rank covariance matrix yields more accurate simulated demand sets while modeling EDP dependence by assuming a joint lognormal distribution may be more efficient, but less accurate overall, especially for lower rise buildings; for higher rise buildings, Gaussian copulas provide more accurate representations of EDPs overall.

In terms of the calculation and assessment of loss, results show that for the majority of cases, losses obtained via EDPs generated using Gaussian copulas better match the loss obtained from the observed, or population, EDPs compared with loss generated from EDPs that follow a joint lognormal distribution. These results shed light on the availability of more statistical tools that require fewer assumptions that can be implemented to generate EDP realization vectors, such as Gaussian copulas. Furthermore, these results highlight the sources of variability that exist in performance based estimations of economic loss at the EDP level, which, in turn, affect the estimation of damage and loss. This study addresses a solution to remove the need for potentially inaccurate assumptions at the EDP level of performance assessment by providing a methodology that can minimize the inaccuracies associated with simulating suites of demand sets from a small number of initial analyses. This study contributes to a more holistically accurate methodology that will allow for more reliable engineered structures.

CHAPTER 4: VALIDATIO OF SIMULATED GROUND MOTIONS VIA VECTOR BASED INTENSITY MEASURE METHOD

4.1 INTRODUCTION

Ground motions selected from historically recorded events have typically been used in nonlinear dynamic analyses for structural response assessment. This process consists of selecting and modifying ground motions to match target seismic hazard characteristics, including response spectrum, magnitude, source-to-site distance and local soil conditions. Although these ground motion selection and modification methods (NIST, 2011; Baker and Cornell, 2006; Luco and Cornell, 2007; Baker, 2011) are widely used in engineering applications, they pose several limitations. Some of these limitations include bias estimates of structural response when using certain ground motion selection and scaling methods (Bazzurro and Luco, 2006; Haselton, 2009) and lack of recorded ground motion data for specific scenarios, namely large-magnitude events at short source-to-site distances (Rezaeian et al, 2015). As an alternative with the potential to overcome these limitations, design codes such as ASCE/SEI 7-16 allow the use of simulated ground motions for engineering applications when the required number of recorded ground motions is not available. Therefore, simulated ground motions are becoming increasingly attractive surrogates to recorded ground motions for performance-based assessment and design of structural systems. With this backdrop, the engineering community is still hesitant in embracing simulated ground motions in practice as they are concerned with how accurate they are in estimating seismic demand of structural systems (Naeim and Graves, 2006; Jones and Zareian, 2010). With significant improvements of simulated ground motions, especially in recent years, the concept of ground motion simulation validation has gained a large amount of attention (Star et al, 2011; Galasso et al, 2012, 2013).

One of the difficulties in the validation of simulated ground motions is the lack of consensus in the engineering research community regarding how to evaluate the sufficiency of simulations for engineering applications. Validation efforts, often times, utilize historical events as the bases of comparison. One major advantage of this is that researchers can directly assess differences in actual recordings and corresponding simulations at the same seismograph stations. For the validation methodology developed in this study, historical records are used for validation with the assumption that if simulation models are sufficient in generating waveforms representative of their recorded counterparts, then they may also be sufficient in predicting future events.

Efforts to validate simulated ground motions entail comparisons of either structural response or ground motion waveform characteristics from simulated and recorded motions. Several studies (Galasso et al, 2012; Atkinson and Goda, 2010; Jayaram and Shome, 2012) have focused on comparing the response of single degree of freedom (SDOF) systems subject to recorded and broadband simulations of historical events. Galasso et al (2012) show the similarities in elastic and inelastic demands for SDOF systems subject to simulated and recorded ground motions while noting that for some systems, differences in median response obtained were dependent on SDOF period, nonlinearly level, and in cases of peak response, spectral shape. Atkinson and Goda (2010) also investigated the response of SDOF systems subject to simulated records along with modified real records and scaled-real records and found that peak nonlinear responses from simulated and modified records are similar. Galasso et al (2013) extended their work from 2012 and compared median seismic response of multi degree of freedom (MDOF) systems subject to simulated and recorded ground motions and found differences in response in the short-period range and higher record-to-record variability of seismic responses.

Besides assessing differences in structural response, other studies (Kristekova et al, 2006; Olsen and Mayhew, 2010) have chosen to focus on validation of waveform parameters through the use of goodness-of-fit measures. The validation methodology developed by Rezaeian et al (2015) utilizes three novel validation metrics, taken from the simulation model of Rezaeian and Der Kiureghian (2010), that are characterized by the evolution of intensity and frequency content of the motion. From these three metrics, easily interpretable parameters are extracted which can then be used to provide clear feedback to simulators regarding how their simulations differ from recorded motions.

Up until this point, previous studies in the area of ground motion simulation validation have assessed differences through (1) response of SDOF and MDOF systems, and (2) ground-motion waveform characteristics. This study attempts to add a third method of validation, that can be seen as a combination of the previous two, which links the differences in waveform characteristics with differences in structural response. This study therefore builds upon and extends the work done in these aforementioned efforts through the development of a ground motions simulation validation methodology that considers how differences in waveform parameters affect differences in structural response of systems subject to simulated and recorded motions. The same metrics utilized by Rezaeian et al (2015) are used for the validation methodology proposed in this paper as they are able to capture nonstationarities in intensity and frequency content of waveforms and have been suggested to influence the nonlinear response of structural and geotechnical systems and. The proposed validation technique is tested on different types of structures (i.e. bridges and buildings) to provide a practical application and understanding of the methodology. Multiple regression analyses are carried out to statistically determine which waveform parameters contribute to predicting response of structures subject to 1994 Mw 6.7 Northridge recorded

earthquake and its simulated counterparts considering four simulation methodologies available via Southern California Earthquake Center (SCEC) Broadband Platform (BBP): 1) stochastic finite-fault (EXSIM) (Motazedian and Atkinson, 2005; Atkinson et al., 2009; Boore, 2009), 2) hybrid simulations by Graves and Pitarka (GP) (Graves and Pitarka, 2014), 3) Irikura recipe (IR) (Irikura and Miyake, 2011), and 4) Song (Song & Somerville, 2010).

4.2 DESCRIPTION OF VBIM METHODOLOGY

The validation methodology proposed in this paper, titled Vector Based Intensity-measure Method (VBIM) ultimately links the differences in waveform characteristics and structural response for simulated and recorded ground motions. The validation methodology is based on the assumption that a ground motion simulation model can acceptably generate waveforms of future seismic events if it is capable to generate the waveforms of past seismic events with agreeable tolerance in key significant ground motion intensity measures (IM). This is accomplished through a step by step procedural analysis where first, significant IMs for both simulated and recorded ground motion pairs are determined through statistical analyses and next, these significant IMs are used to predict future estimates of structural response. The four IMs considered in this study, taken from the simulation model of Rezaeian and Der Kiureghian (2010), are Arias intensity (I_a), effective duration of the motion ($D_{5\text{-}95}$), filter frequency (ω_{mid}) at the time in which 45% of I_a is reached, and the rate of change of the filter frequency with time (ω'). Second, differences in the simulated and recorded prediction models developed in step 1 are assessed and third, the sensitivity of structural demand to changes in these specific waveform parameters is quantified.

STEP 1: Identification of significant IMs in predicting EDP through regression analysis

Regression models are established using engineering demand parameters (EDPs) and IMs of both recorded and simulated ground motions and goodness-of-fit measures are used to assess how well these models fit the sample data. In terms of ground motion simulation validation, regression analysis can be used to determine the metrics (i.e. IMs) that are statistically significant in predicting response of structures (i.e. EDPs) subject to simulated and recorded ground motions. The general regression equation for recorded/simulated data is shown by (4-1) [(4-2)], where EDP^r [EDP^s] is the engineering demand parameter of recorded [simulated] ground motions, IM_i^r [IM_i^s] is the intensity measure of recorded [simulated] ground motions, a^r [a^s] is the intercept of the recorded [simulated] regression line, a_i^r [a_i^s] is the slope of the i^{th} IM (total of N IMs) in the recorded [simulated] regression model, and ε^r [ε^s] is the random variance of the recorded [simulated] regression model.

$$\ln(EDP^r) = a^r + \sum_{i=1}^{N} a_i^r \ln(IM_i^r) + \varepsilon^r \quad (4\text{-}1)$$

$$\ln(EDP^s) = a^s + \sum_{i=1}^{N} a_i^s \ln(IM_i^s) + \varepsilon^s \quad (4\text{-}2)$$

Therefore, from regression analyses, we can determine the following information: (1) significant recorded and simulated IMs in predicting EDPs from recorded and simulated ground motions, respectively, (2) recorded and simulated models that predict EDP based on significant IMs and (3) the goodness-of-fit of these regression models (R^2 and standard error).

STEP 2: Comparison of simulated and recorded regression models used to estimate EDP from significant IMs

An Analysis of Variance (ANOVA) test (Girden, 1992) is conducted in order to determine if there is a significant statistical difference between the simulated and recorded regression models developed in Step 1. ANOVA is a statistical tool that, through a series of calculations, can shed light on the level of variability within and between regression models (Lacey, 2018). In order to test a hypothesis that detects differences in two regression models using ANOVA, the models being tested must be nested, or hierarchical, in that one only differs from the other by the addition of a single extra independent variable. As regression equations (1) and (2) are not nested models, a typical ANOVA that compares two nested equations cannot be used. Instead, the models are combined through the addition of an indicator variable in order to test the null hypothesis, H_0: $a_i^r = a_i^s$, against the alternative hypothesis, H_a: $a_i^r \neq a_i^s$. An indicator variable, sim_i, is added into the data, coded as 0 for recorded and 1 for simulated ground motions, and another variable, $sim_i * IM_i$, is added which is the product of sim_i and the IM(s) included in the regression model. In the new regression model, (4-3), sim_i, IM_i and $sim_i * IM_i$ are used as the independent variables and ANOVA is then used to test the null hypothesis.

$$EDP = a^r + \sum_{i=1}^{N}[(a_i^s - a_i^r)(sim_i) + (a_i^r)(IM_i) + (a_i^s - a_i^r)(IM_i)(sim_i)] \quad (4\text{-}3)$$

From ANOVA, an F test statistic, which is the ratio of between-groups variance to within-groups variance, is determined. In order to calculate the F statistic, the total variance in the dataset must first be estimated through

$$s^2 = \frac{\sum_{i=1}^{N}(EDP_i - \overline{EDP})^2}{n-1} \quad (4\text{-}4)$$

where s is the standard deviation (also referred to as the total mean square, MS_T) of the data and n is the number of observations. For ANOVA, the numerator of (4-4) is referred to as the total sum of squares (SS_T) and the denominator represents the degrees of freedom of SS_T. The model sum of squares, SS_R, can be found in a similar way as SS_T, but estimates of EDP, $\widehat{EDP}$, are substituted for true observations of EDP.

$$SS_R = \sum_{i=1}^{N}\left(\widehat{EDP_i} - \overline{EDP}\right)^2 \quad (4\text{-}5)$$

The corresponding model mean square, MS_R, with degrees of freedom $dof(SS_R)$, is $MS_R = SS_R/dof(SS_R)$. Now, the residual error, e_i, which is the difference between the actual observations and the model estimates, can be used to determine the error mean square, $MS_E = SS_E/dof(SS_E)$. The total sum of squares, SS_T, which explains the total variability in the observed data, is the sum of SS_R and SS_E and the F statistics, F_0, is MS_R/MS_E. Once the F statistic is determined, a p-value is assigned which can be tested against the significance level. In this step, a significance level of 5% is used. If the p-value is less than 5%, the null hypothesis is rejected and $a_i^r \neq a_i^s$.

STEP 3: Quantification of uncertainty in future predictions of EDP based on significant IMs

VBIM assumes that waveform parameters and responses of recorded motions are the direct observations, or ground truth, and that simulated ground motions developed based on specific historical records should have similar properties. Hence, a direct comparison of the mathematical difference between simulated and recorded ground motion properties is justified. This direct comparison is accomplished through the calculation of IM_i^s/IM_i^r plotted against EDP_i^s/EDP_i^r,

thus, quantifying the differences in EDP that correspond to differences in IM. Next, predictive inferences can be made regarding future differences in EDP from simulated and recorded ground motions based on the observed data through confidence and prediction intervals. A confidence interval is used to estimate future values of mean response at a certain value of $IM_i = IM_i^*$ given input parameters such that

$$\hat{\mu}_{EDP|IM_i^*} = a + \sum_{i-1}^{N}(a_i {IM_i}^*) \tag{4-6}$$

where IM_i^* is the mean for a given subpopulation, *a* is the intercept of the regression equation and a_i is the slope of IM_i^* . The uncertainty in parameters *a* and a_i are

$$SE_{\mu_{EDP|IM_i^*}} = s\sqrt{\frac{1}{n} + \frac{\left(IM_i^* - \overline{IM_i}\right)^2}{\sum\left(IM_i^* - \overline{IM_i}\right)^2}} \tag{4-7}$$

where *n* is the number of sample points, $\overline{IM_i}$ is the sample mean and *s* is the standard deviation of the sample. The confidence interval of a sample (using a t-statistic) is represented as

$$\hat{\mu}_{EDP|IM_i^*} \pm t_{1-\alpha/2,n-2} SE_{\hat{\mu}_{EDP|IM_i^*}} \tag{4-8}$$

where $t_{1-\alpha,n-2}$ is the t-statistic at confidence level of *1-α.* Similar to a confidence interval, a prediction interval can be used to estimate future observations of structural response, $\widehat{EDP}$, at a certain value of $IM_i = IM_i^*$ such that

$$\widehat{EDP}(IM_i^*) = a + \sum_{i=1}^{N}(a_i IM_i^*) \tag{4-9}$$

where the uncertainty in a and a_i are measured by

$$SE^2_{\widehat{EDP}(IM_i^*)} = \sum_{i=1}^{N} s^2_{EDP|IM_i} + SE^2_{\hat{\mu}_{EDP|IM_i^*}} \ , \ s^2_{EDP|IM_i} = \frac{1}{n-2}\sum_{i=1}^{N} e_i^2 \qquad (4\text{-}10)$$

$$\widehat{EDP}(IM^*) \pm t_{1-\alpha/2,n-2} SE_{\widehat{EDP}(IM^*)} \qquad (4\text{-}11)$$

and e_i is the deviation of the sample. With the addition of an extra parameter that represents deviation of the sample, the prediction interval is wider than the confidence interval. Confidence and prediction intervals can be calculated for the developed regression models, which provide ranges of future predictions of structural response based on the significant intensity measures.

4.3 APPLICATION OF VBIM ON CASE STUDY STRUCTURES

4.3.1 Description of Case Study Structures and Demand Parameters

VBIM is illustrated through the use of three case study structures including two special steel moment resisting frame (SMRF) buildings and one concrete bridge. The SMRF buildings are designed based on ASCE/SEI 7-02 (ASCE, 2005) and ANSI/AISC 341-05 (AISC, 2005) for a site in downtown Los Angeles, California with typical soil site class D. Interstory drift ratios (IDRs) for these buildings were used as the EDPs representative of structural response. Nonlinear response history analysis was conducted to obtain the values of IDR when the buildings were subjected to the 1994 M_w 6.7 Northridge record and the corresponding simulations obtained via the Southern California Earthquake Center (SCEC) Broadband Platform (BBP). Figures 5.1 and 5.2 represents the plan and elevation views for these case study buildings and Figure 5.3 displays the pushover analyses.

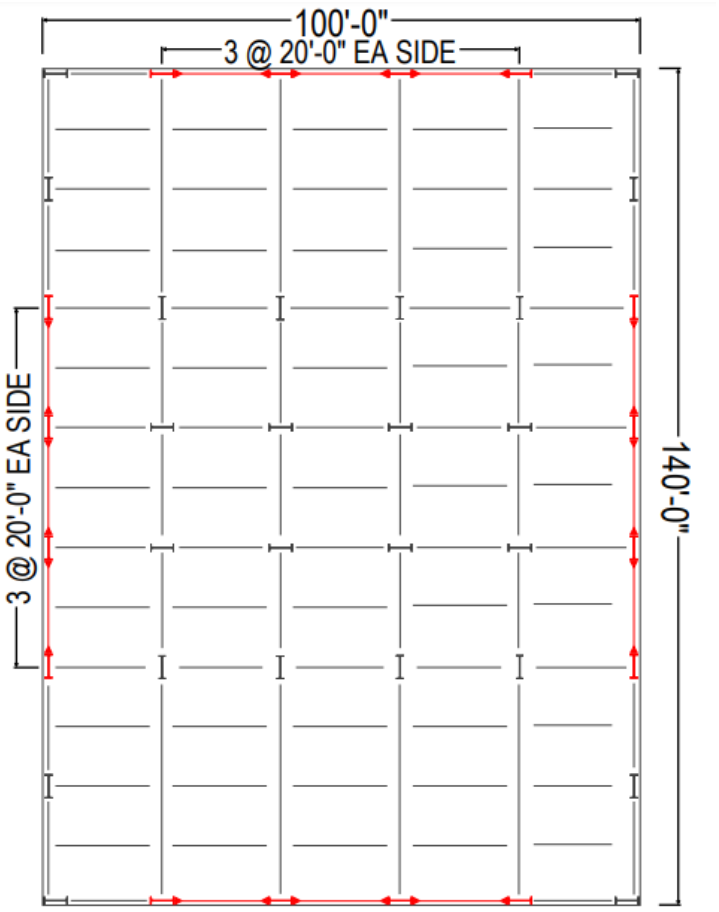
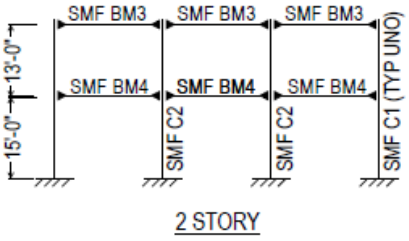


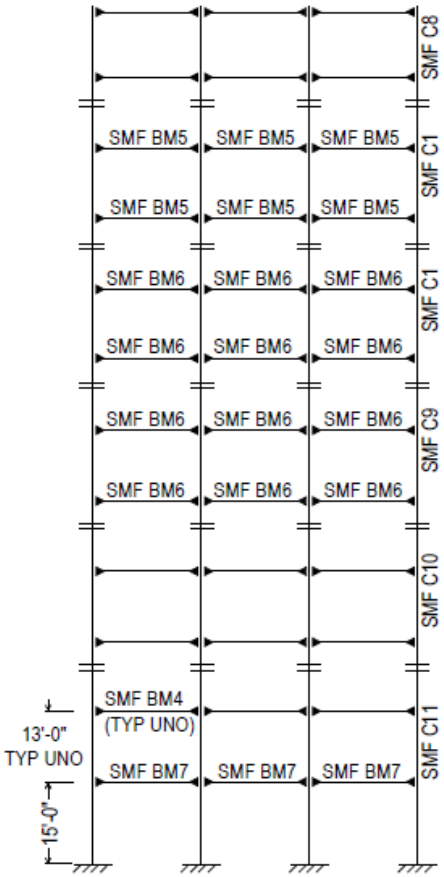
Figure 4.1: Plan view for 2- and 12-story SMRF buildings

SMF BEAM AND COLUMN SCHEDULE			
BEAMS		COLUMNS	
LABEL	SIZE	LABEL	SIZE
SMF BM3	W16X31	SMF C1	W24X31
SMF BM4	W30x132	SMF C2	W24x62
SMF BM5	W27x94	SMF C8	W24x84
SMF BM6	W30x116	SMF C9	W24x146
SMF BM7	W30x124	SMF C10	W24x162
		SMF C11	W24x207

NOTE:
ALL COLUMN SIZES NOT INDICATED ON ELEVATION SHALL CONFORM TO THE SIZE INDICATED FOR THAT LEVEL



2 STORY



12 STORY

Figure 4.2: Plan view for 2- and 12-story SMRF buildings

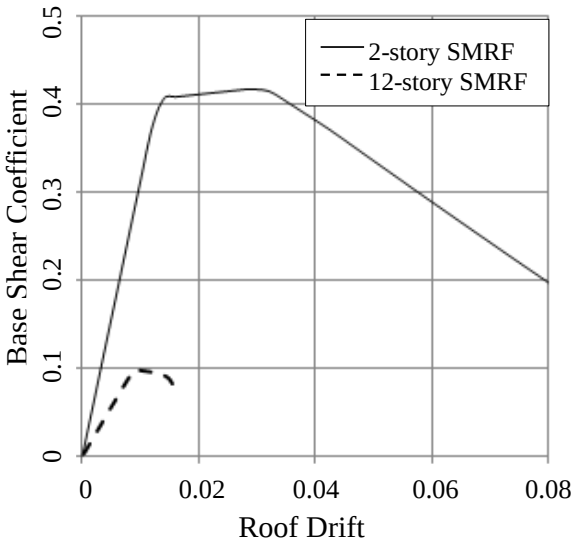


Figure 4.3: Pushover curves for 2- and 12- story SMRFs

VBIM is also applied to the Jack Tone Road On-Ramp Overcrossing built in 2001. This two-span, single-column bridge was modeled by Kaviani et al (2012) following recommendations provided by Caltrans Seismic Design Criteria Version 1.6 (Caltrans, 2010) and research conducted by Aviram et al. (2008) via OpenSees (McKenna et al., 2000). The bridge has two spans with one 22-ft. tall reinforced concrete column supporting two equal 110.25 foot span cast-in-place concrete decks. The deck is modeled with elastic beam-column elements. The superstructure of the bridge is a three-cell continuous pre-stressed reinforced concrete box girder. . Deep foundation support is provided to the bent column through a group of 25 driven H-piles at 36 feet in height while the abutment skew angle is taken as 33˚.The deck of the highway bridge consists of cast-in-place concrete and has a width of 27.1 feet and depth is 4.64 feet. The single bent column supporting the two deck spans has a diameter of 5.51 feet. This bridge is part of a group of bridges that are modeled with characteristics representative of common bridges in California and is located in

Southern California, a high seismicity region. The point of interest on the bridge to collect seismic response (i.e., EDP) is at the mid height of the cross-section of the deck of the bridge, vertically aligned with the column of the bridge. The displacement in the longitudinal and transverse direction was recorded and the square root of the sum of the squares (SRSS) of the longitudinal and transverse records was taken as the EDP of the bridge. The SRSS of each waveform parameter was determined through (4-12), (4-13), and (4-14) for I_a, $D_{5\text{-}95\%}$, and ω_{mid} respectively.

$$I_a = \sqrt{I_{ax}^2 + I_{ay}^2} \tag{4-12}$$

$$D_{5-95\%} = \sqrt{D_{5-95\%\,(x)} D_{5-95\%\,(y)}} \tag{4-13}$$

$$\omega_{mid} = \sqrt{\omega_{midx} \omega_{midy}} \tag{4-14}$$

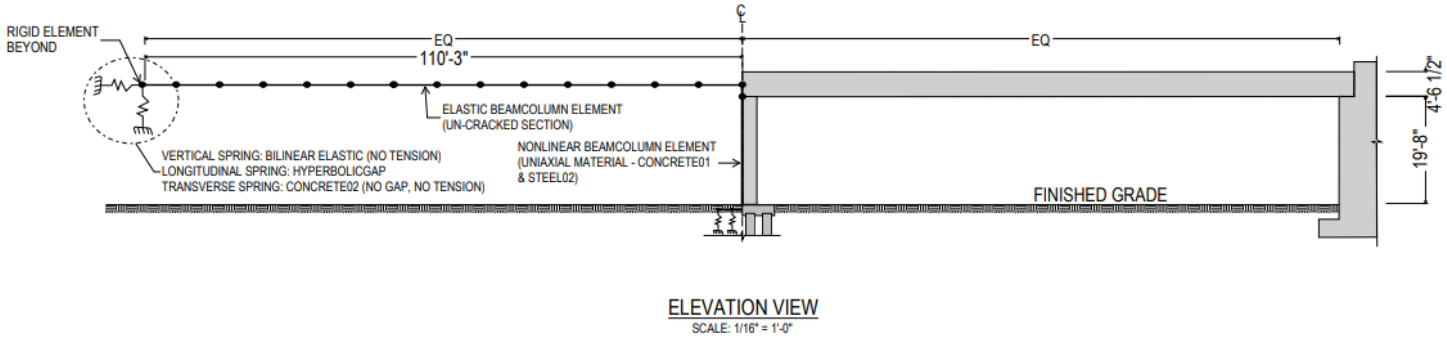


Figure 4.4: Details of case study bridge model

4.3.2 VBIM APPLICATION AND RESUTLS

Significant IMs in predicting EDP

The EDPs of the case study structures subject to both simulated and recorded GMs were determined via NDLA and then used as the dependent variables in (4-15) and (4-16), respectively. Several regression analyses were conducted

$$\ln(EDP^r) = a^r + a_1^r \ln(I_a^r) + a_2^r \ln(D_{5-95\%}^r) + a_3^r \ln(\omega_{mid}^r) + a_4^r \ln(\omega'^r) + \varepsilon^r \quad (4\text{-}15)$$

$$\ln(EDP^s) = a^s + a_1^s \ln(I_a^s) + a_2^s \ln(D_{5-95\%}^s) + a_3^s \ln(\omega_{mid}^s) + a_4^s \ln(\omega'^s) + \varepsilon^s \quad (4\text{-}16)$$

Equation (5-15) is used as the base regression model for comparison, as it is representative of the historical ground motion. 15 recorded and 48 simulated regression models are developed based on (5-15) and (4-5), respectively. These models correspond to the following conditions with 4 considered simulation methodologies (EXSIM, GP, Song, Irikura-Recipe):

1. Bridge: 1 recorded regression model and 4 simulated regression models for a total of 5 regression models
2. 2-story SMRF: 1 recorded regression model per floor (2) and 4 simulated regression models per floor (8) for a total of 10 regression models
3. 12-story SMRF: 1 recorded regression model per floor (12) and 4 simulated regression models per floor (36) for a total of 48 regression models

For all 15 recorded and 48 simulated (63 total) regression models, I_a, is statistically significant in predicting EDP. The differences between the bridge and building systems arise in terms of the subsequent IMs that were considered. For both buildings systems (58 total models), I_a remains to

be the only significant IM in the regression model whereas for the bridge system (5 total models), along with I_a, $D_{5\text{-}95\%}$ is significant. These findings are consistent for both historical Northridge records and the simulated counterparts considering all four simulation methodologies. Although these four IMs have been proven to be representative of the evolution of intensity and frequency content of ground motions, this study is the first to link these parameters to the response of structural systems. This step of the validation procedure accomplishes this task.

Goodness of fit measures for recorded and simulated regression models

Tables 5.1, 5.2 and 5.3 display the R^2 and standard error for the buildings and Table 5.4 displays the R^2 and standard error values for the bridge. There are several noteworthy conclusions that can be drawn from these goodness-of-fit metrics:

(1) There is a significant difference between the statistics of the regression models for the buildings and the bridge. For the bridge system, the waveform parameters are able to explain a large amount of the EDP (R^2 is around 90% for all models) whereas for both buildings, the waveform parameters are not able to explain nearly as much (R^2 is around 45%). There is a general consensus in several statistical applications that an R^2 value of 50% or greater constitutes an acceptable regression model (Cameron, 1997). Therefore, the bridge waveform parameters are highly representative of the EDP. For the buildings, the R^2 values, specifically for the recorded model, are very close to 50%, and some simulation methodologies (Song for the 2-story and Irikura-Recipe for the 12-story) are able to surpass the 50% threshold. This implies that the models are good enough to predict the EDP to some extent, with potential room for improvement with the addition of other parameters that may be predictive of structural response.

(2) Table 5.1 and highlight the similarities between simulated and recorded data in terms of how well the waveform parameters from simulated and recorded ground motions predict structural response. The maximum difference in standard error for the 2-story model occurs when using GP simulation methodology, with an error of 31%, and the maximum difference in R^2 for the 2-story building occurs when using EXSIM simulation methodology, with an error of 21%. The maximum error in standard error for the bridge model is 16% for both GP and Irikura-Recipe, about half that of the building, and the maximum error for R^2 it is 4% for EXSIM. The results of the bridge highlight the overall similarities in the goodness-of-fit between simulated and recorded models while for the building, there is room for improvement in how well the simulated models match their recorded counterparts.

(3) For the 12-story model, a significant difference can be seen in several instances between simulated and recorded goodness of fit measures. For example, the largest error in standard error (56%) occurs when using EXSIM simulation methodology at the sixth story, and the largest error in R^2 (22%) occurs also when using EXSIM but for the seventh story. These are the worst case scenarios, with several of the other goodness-of-fit measures measuring very closely for simulated and recorded data (for example, R^2 values for GP data are similar to recorded data for all stories).

Table 4.1: Standard Errors and R-squared values from multiple regression analysis of 2-story SMRF

	2-Story Model			
	Standard Error		R^2	
	Floor 1	Floor 2	Floor 1	Floor 2
REC	0.72	0.74	0.47	0.47
EXSIM	0.52	0.53	0.34	0.37
GP	0.52	0.51	0.39	0.4
SONG	0.65	0.64	0.52	0.52
IRIK	0.55	0.54	0.38	0.41

Table 4.2: R-squared values from multiple regression analysis of 12-story SMRF

R^2: 12-story model												
	Floor 1	Floor 2	Floor 3	Floor 4	Floor 5	Floor 6	Floor 7	Floor 8	Floor 9	Floor 10	Floor 11	Floor 12
REC	0.46	0.46	0.45	0.46	0.46	0.45	0.46	0.44	0.44	0.44	0.41	0.41
EXSIM	0.37	0.35	0.33	0.32	0.35	0.35	0.35	0.37	0.38	0.4	0.44	0.47
GP	0.43	0.44	0.44	0.44	0.43	0.43	0.42	0.43	0.42	0.43	0.43	0.48
SONG	0.49	0.48	0.48	0.44	0.43	0.43	0.43	0.43	0.43	0.45	0.47	0.49
IRIK	0.49	0.48	0.49	0.5	0.51	0.53	0.5	0.5	0.46	0.45	0.47	0.49

Table 4.3: Standard errors from multiple regression analysis of 12-story SMRF

Standard Error: 12-story model												
	Floor 1	Floor 2	Floor 3	Floor 4	Floor 5	Floor 6	Floor 7	Floor 8	Floor 9	Floor 10	Floor 11	Floor 12
REC	0.66	0.71	0.72	0.71	0.72	0.75	0.79	0.81	0.82	0.82	0.87	0.93
EXSIM	0.52	0.49	0.51	0.52	0.52	0.51	0.35	0.48	0.48	0.48	0.46	0.45
GP	0.62	0.58	0.59	0.62	0.61	0.43	0.58	0.59	0.61	0.58	0.52	0.48
SONG	0.89	0.83	0.77	0.76	0.76	0.74	0.76	0.76	0.77	0.71	0.65	0.58
IRIK	0.52	0.52	0.54	0.55	0.57	0.53	0.56	0.52	0.54	0.45	0.53	0.51

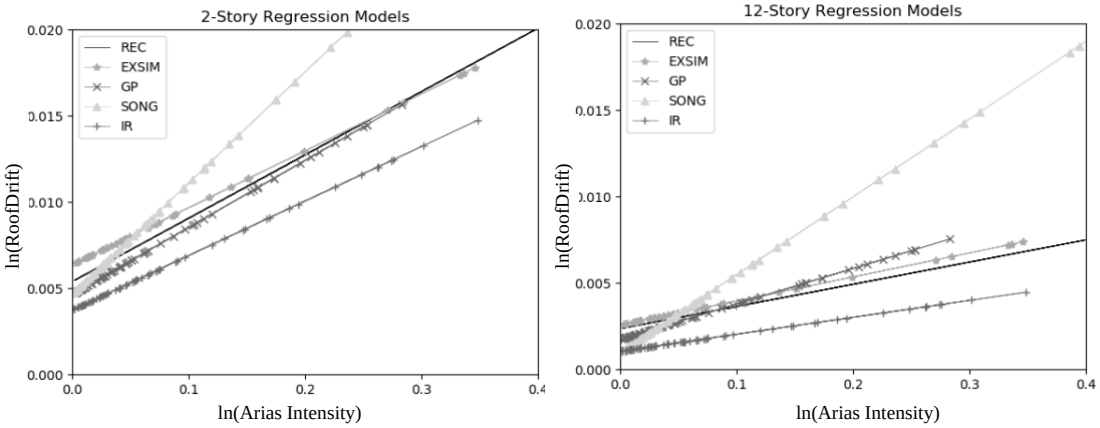
Table 4.4: Standard Errors and R-squared values from multiple regression analysis of bridge

Bridge Model- Combined Orthogonal Components		
	Standard Error	**R^2**
REC	0.31	0.93
EXSIM	0.26	0.89
GP	0.23	0.92
SONG	0.25	0.91
IRIK	0.23	0.91

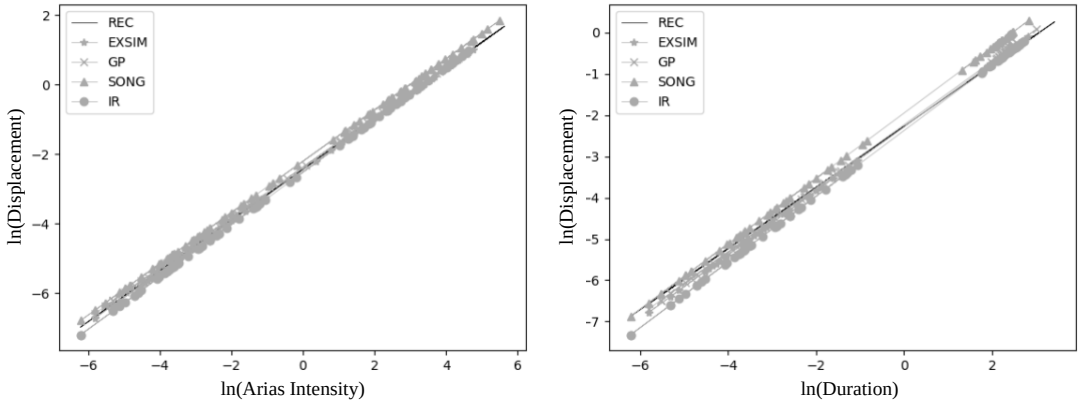
To investigate whether there is a statistical difference between recorded and simulated regression models, ANOVA is conducted. For all ANOVA tests (2-story building, 12-story building, and concrete bridge), the $H_{0,}$ —which states that regression coefficients for simulated and recorded models are equal—cannot be rejected as the p-value (shown in Table 5.5) is greater than the chosen significance level of 0.05. Given failure to reject the null hypothesis for all models, it can be argued that the models developed using waveform parameters to predict structural response are statistically similar for simulated and recorded data. Figure 4 shows the trendlines of the each recorded and simulated regression model and provides a graphical representation of the similarity between these models, reinforcing the results from ANOVA. Although there is much less scatter for the bridge system compared with both buildings, the buildings still share similar slopes and intercepts between recorded and simulated motions, especially in the case of the 2-story building.

Table 4. 5: P-values from ANOVA for case study structures with significance level 0.05

	2-story	12-story	Bridge
EXSIM	0.63	0.83	0.89
GP	0.82	0.21	0.72
SONG	0.07	0.08	0.15
IRIK	0.52	0.57	0.66



(a) (b)



(c) (d)

Figure 4.5: Trendlines of regression models for simulated ground motions compared with recorded regression models shown with a black solid line for (a) the bridge considering Ia, (b) the bridge considering D5-95% , and the SMRF structures considering Ia

Sensitivity of EDPs to differences in select IMs of recorded and simulated ground motions

Results of ANOVA show that the statistical models developed from simulated and recorded data to predict response based on select waveform parameters are similar. The next step of the validation aims to investigate how differences in significant waveform parameters affect differences in structural response. Shown in Figures 5.5 and 5.6 are the log ratio comparisons of significant IMs plotted against structural response for the cases study structures, along with a 95% confidence interval (grey shaded area) and prediction interval (are within the dashed red lines). For the buildings, there is a vertical clustering around 0 on the plot, which indicates that I_a is generally similar for simulated and recorded ground motions. However, this is not always the case for response as there is a large range of variability in response, even when I_a of simulated and recorded motions match well. Differences in bridge response (Figure 5.6) are highly correlated with the differences in waveform parameters, especially I_a. These results indicate that the considered waveform parameters, I_a in particular, are strong predictors of bridge response of both simulated and recorded motions and that the differences in these parameters directly correlate with differences in bridge response. In other words, when recorded/simulated IMs are overestimated or underestimated, recorded/simulated EDPs are also overestimated or underestimated, respectively. However, although the waveform parameters of simulated and recorded motions are similar, there may be additional parameters that are more representative of the difference between building response from simulated and recorded ground motions.

The sensitivity of response to changes in significant waveform parameters is investigated through the determination of confidence and prediction intervals. Larger margins for both confidence and prediction intervals can be seen when comparing building and bridge results, which further shows the better predictive accuracy of waveform parameters for bridge response. The

large prediction interval for buildings suggests that there is a wide range of possibilities that future predictions may lie. In order to quantify these differences in IMs and EDPs, a 5% margin of error in the estimation of structural response between recorded and simulated ground motions is considered. A 5% margin of error between the structural response of recorded ground motions and their simulated counter-parts is calculated by $(EDP^s/EDP^r) = 1.05$. Taking the natural log of both sides of this equation yields $ln(EDP^s/EDP^r) = ln(1.05)$, where $ln(1.05)$ is equal to approximately 0.05. Therefore, a log ratio of 0.05 below or above the 0 mark of the y-axis on Figures 5.5 and 5.6 corresponds to a 5% difference between simulated and recorded motions. For building response, the majority of IM ratios lie within approximately ± 0.5 (corresponding to a log ratio of about ± 0.5), however, in order to achieve 5% accuracy in the estimation of response, the I_a of simulated ground motions must fall within about ± 0.25 from or, in other words, when the log ratio of simulated to recorded IMs is approximately ± 0.25. For bridge response, there is a much stronger correlation how well the simulations IMs match the recorded IMs and the accurate prediction of response. The majority of IM ratios fall within the acceptable range, suggesting that these simulations are able to match their recorded counterparts well.

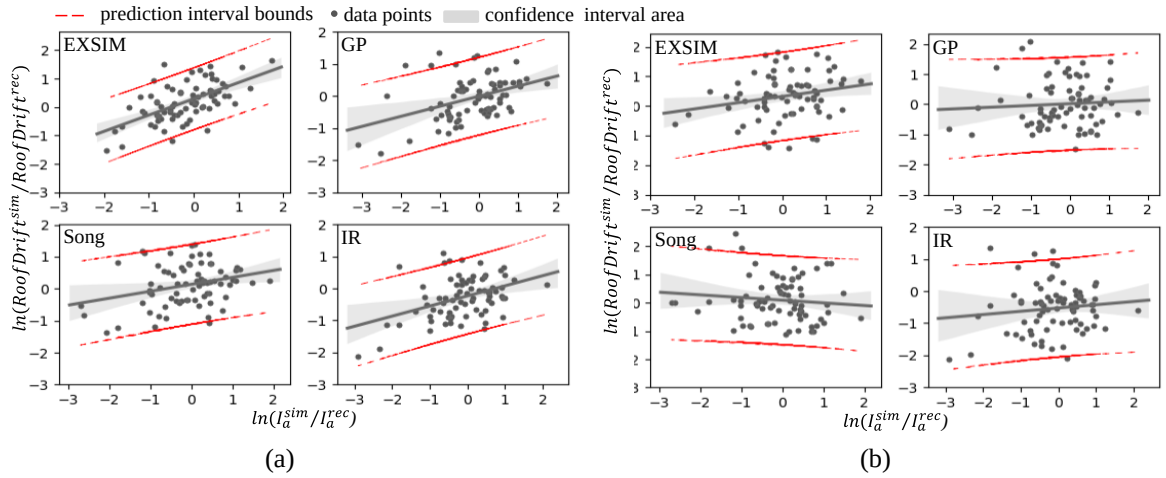


Figure 4. 6: Ratio plot of simulated and recorded regression models for the prediction of roof drift based on Arias Intensity with 95% confidence interval band represented by the grey shaded area around the line-of-best fit and prediction interval represented by the red dashed lines

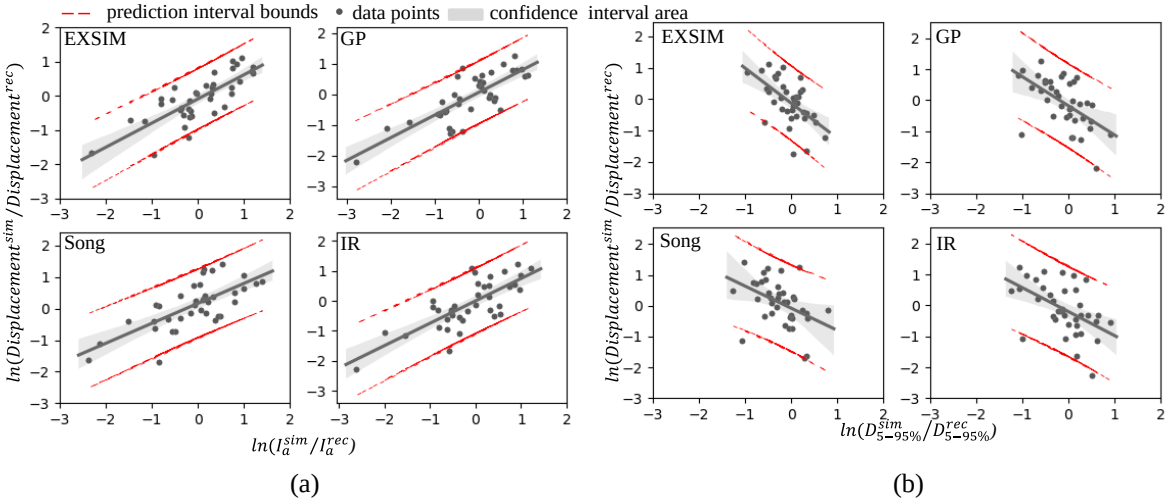


Figure 4.7: Ratio plot of simulated and recorded regression models for the prediction of displacement based on (a) Arias Intensity and (b) significant duration with 95% confidence interval band for concrete bridge structure with the confidence interval represented

Secondary verification of significant metrics in predicting structural response through variable transformation into the standard normal space

When conducting regression analyses with multiple variables to determine which parameters are significant in predicting structural response, it is important to consider how the individual statistical distributions of these variables affects the outcomes of regression. Razaeian and Der Kiureghian (2010) detailed the unique individual distributions of each of these waveform metrics, which were considered in this study as an additional validation procedure to assess whether transforming the variables into the standard normal space would affect how well they predict structural response. Transformations into the standard normal domain are applied according to methodology discussed in PEER Report 2010/02 (Rezaeian and Der Kiureghian, 2010). Each intensity measure, along with corresponding distributions, is represented as follows: ln(Ia)—Standard Normal Distribution, $D_{5\text{-}95}$—Beta Distribution, ω_{mid} —Gamma Distribution, ω'—Two-sided Truncated Exponential Distribution. For all case study structures, transforming the variables into the standard normal space and then carrying out the regression analyses did not alter the previous results of this study. I_a was still shown to be the only waveform metric significant in predicting building response while I_a, $D_{5\text{-}95\%}$ were the only significant waveform metrics in predicting bridge response.

VBIM application for IMs that do not show significance in regression analyses

The validation procedure outlined in VBIM consists of a process of elimination in regards to IMs, where after the first step of validation, only IMs that are statistically significant in predicting EDP are included in the remainder of the validation methodology. For example, for the 2- and 12-story SMRFs, only I_a was shown to be significant in predicting the EDP, and therefore only this IM was included in the calculation of confidence and prediction intervals. Here, VBIM is applied in one scenario for an IM (ω' for the 2-story SMRF) that did not show significance in predicting EDP in step 1. Figure 5.8 presents the natural log ratio of simulated and recorded data corresponding to the 2-story SMRF for ω'. Comparing Figure 8 (based on non-significant IM, ω') and Figure 5.6 (based on significant IM, I_a), several observations are drawn:

(1) The significantly larger prediction intervals in Figure 5.8 signify that future point estimates of roof drift based on ω' have a much larger window of possibility compared with those based on I_a.

(2) Clear slopes can be seen for the lines-of-best-fit presented in Figure 5.6, showing the dependency of EDP on I_a whereas the lines-of-best-fit in Figure 5.8, along with the prediction and confidence interval bounds, are almost perfectly horizontal. Therefore, it can be concluded that regardless of how similar values of ω' are for simulated and recorded ground motions, observed values of EDP, along with future predictions of point and mean estimates, remain unaffected.

(3) Figure 5.6 and Figure 5.8 both show data points clustered around 0 on the x-axis, but unlike Figure 5.6, the data in Figure 5.8 is not centered around 0 on the y-axis. Therefore, this shows

that values of ω' between simulated and recorded ground motions may be similar, but this similarity has no effect on how closely the EDP of simulated and recorded motions match.

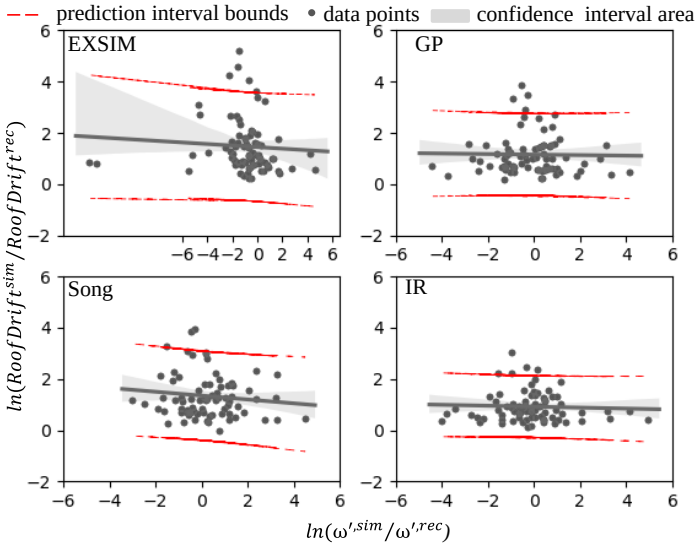


Figure 4.8: Ratio plot of simulated and recorded regression models for the prediction of roof drift based on ω' with 95% confidence interval and prediction interval for 2-story SMRF

VBIM application using spectral acceleration as an IM

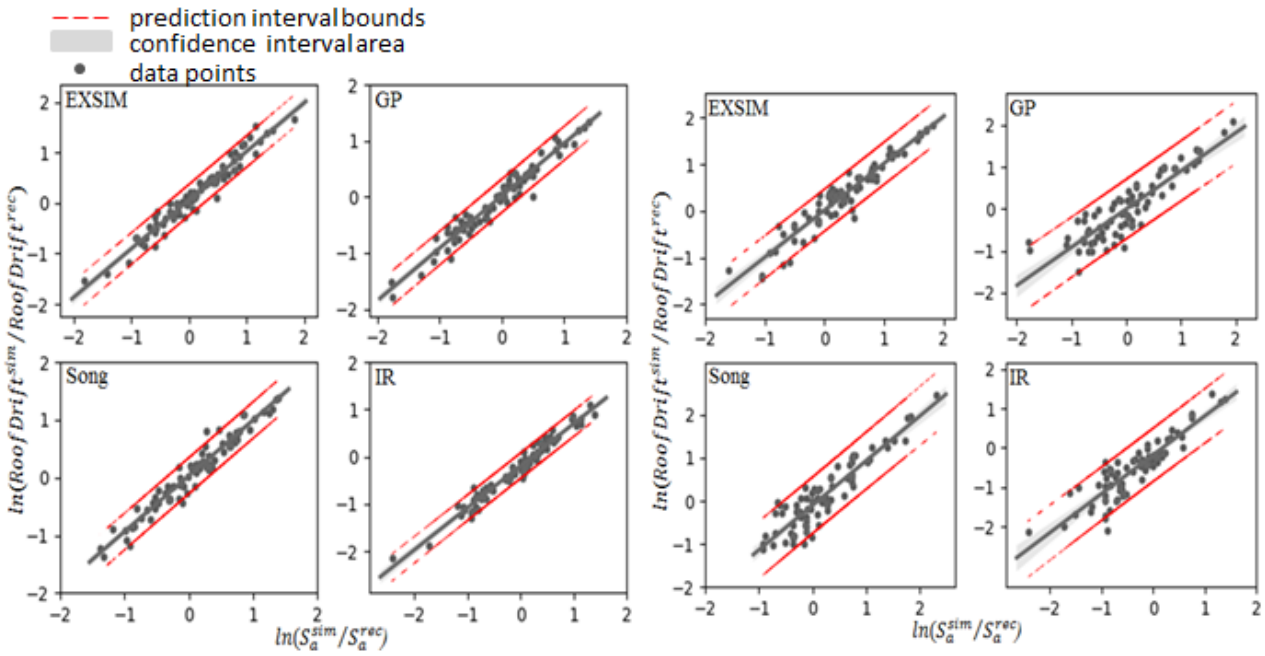
In addition to applying VBIM using both significant and non-significant ground motion waveform parameters as IMs, VBIM is also demonstrated with the addition of spectral acceleration (S_a) as an IM. S_a is chosen for several reasons including its wide used in engineering applications, its direct correlation with the response of structures, as it is a structural period-dependent IM, and its use in a wide range of previous ground motion simulation validation studies as a useful metric for validation. The regression equation with the inclusion of S_a is shown by Equation (4-16). Prediction and confidence intervals (step 3 of VBIM) for the validation methodology considering S_a is presented in Figure 5.7. The strong correlation between accurate estimates of S_a and accurate estimates of response can be seen through the trendlines of the plotted data. A general clustering

around 0 (for both the x- and y-axis) signifies that these simulation methodologies are generally accurate in terms of S_a and correspondingly, in roof drift. Of the four presented simulation methodologies, EXSIM has the narrowest prediction interval, and therefore, estimates of future response for EXSIM are expected to match the response of corresponding recorded ground motions well if values of S_a from EXSIM and recorded ground motions are similar. Consider a 5% difference between simulated and recorded values of S_a. For EXSIM, the prediction interval around $\pm$5% S_a is between approximately -1.5 and 0.04; taking the exponential of these values leads to 0.22 and 1.04. Therefore, if simulated values of S_a are within $\pm$5% of recorded S_a, future prediction of deformation response from ground motions simulated using EXSIM are expected to be between 4% (exponential of 0.04) and 22% (exponential of -1.5) of response from corresponding recorded ground motions. Compared to the predictions based on ω', a non-significant parameter in predicting the structural response of the 2-story SMRF, S_a yields much more accurate estimates of future deformation response.

$$\ln(EDP) = a + a_1\ln(I_a) + a_2\ln(D_{5-95}) + a_3\ln(\omega_{mid}) + a_4\ln(\omega') + a_5\ln(S_a) + \varepsilon_i \quad (4\text{-}16)$$

Table 4.6: Standard Errors and R2 values from multiple nonlinear regression analysis of 2- and 12-story buildings including Sa

	2-Story Model				12-Story Model	
	Standard Error		R^2		Standard Error	R^2
	Floor 1	Floor 2	Floor 1	Floor 2	average of all stories	
REC	0.11	0.11	0.99	0.99	0.39	0.85
EXSIM	0.09	0.09	0.98	0.98	0.28	0.78
GP	0.11	0.11	0.97	0.97	0.31	0.82
SONG	0.12	0.11	0.98	0.98	0.38	0.83
IRIK	0.11	0.1	0.98	0.99	0.29	0.84



(a) (b)

Figure 4.9: Ratio plot of simulated and recorded regression models for the prediction of roof drift based on spectral acceleration with 95% confidence interval band represented by the grey shaded area around the line-of-best fit and prediction interval represented

From the regression analyses, S_a was shown to be a strong predictor of the structural response of buildings. Confidence and prediction intervals of the log ratios of S_a of simulated to recorded motion are determined. Figure 5.7 shows the strongly correlated relationship between accurate estimations of S_a for simulated ground motions and accurate values of roof drift. This clear linear relationship leads to the conclusion that there is a much smaller interval of acceptance when it comes to accurately matching S_a from simulated motions with their recorded counterparts because of the strong sensitivity of response to the accurate estimation of S_a. It is worth noting that for all waveform log ratio plots the majority of points are clustered around 0, which leads to the conclusion that the simulations are, for the most part, accurately representing the recorded motions at the response level, even when the simulated IMs of interest may not match recorded IMs.

4.4 CONCLUSIONS

The validation of simulated ground motions has been the focus of several studies in recent years which have led to the increased confidence in their use. This study aims to add to these efforts by introducing a new methodology for the validation of simulated ground motions titled Vector Based Intensity Measure method (VBIM), which combines significant intensity measures that describe the waveform parameters of ground motions and quantifies how well they are able to measure structural response for recorded and simulated ground motions. Using three case study structures, multiple regression analyses are performed to 1) identify the vector of intensity measures that is able to best predict structural response, 2) develop models that predict structural response from both recorded and simulated ground motions, and 3) quantify similarities and differences between simulated and recorded ground motions using the developed models. Results for the 2- and 12-story SMRFs show that the only significant waveform parameter in predicting roof drift is Arias Intensity; for the bridge structure, both Arias intensity and strong motion duration

are significant. The regression models (i.e., one based on simulated ground motions and another with recorded ground motions) that estimate structure response to vector of intensity measures are statistically compared through ANOVA. Results of ANOVA show no significant difference between regression models of recorded and simulated ground motions out of the four considered simulation methodologies (i.e. Graves and Pitarka, EXSIM, Song, Irikura-Recipe). Through the determination of confidence and prediction intervals of the response based on significant intensity measures, the acceptable range (i.e., tolerance) required in intensity measures in order for simulated ground motions to accurately predict values of response is determined. When considering Arias intensity, there is a much larger interval that encompasses differences in future predictions of building response compared with bridge response for simulated and recorded ground motions.

The results of this study shed light on the similarities between simulated ground motions and their recorded counterparts at the ground motion waveform level and at the structural response level. Furthermore, these results can be used to provide ground motion simulation modelers with recommendations regarding the level of accuracy of intensity measures that is needed in order for future estimates of structural response from simulated ground motions to be within an acceptable range of the structural response from recorded ground motions.

4.5 REFERENCES

American Institute of Steel Construction (AISC) (2005). 360-Specification for Structural Steel Buildings.

American Society of Civil Engineering (ASCE) (2010). Minimum design loads for buildings and other structures (7–10), Standards ASCE/SEI 7–10.

Atkinson, G. M., & Assatourians, K. (2014). Implementation and validation of EXSIM (a stochastic finite-fault ground-motion simulation algorithm) on the SCEC broadband platform. *Seismological Research Letters*, *86*(1), 48-60.

Atkinson, G. M., Ghofrani, H., & Assatourians, K. (2015). Impact of induced seismicity on the evaluation of seismic hazard: Some preliminary considerations. *Seismological Research Letters*, *86*(3), 1009-1021.

Aviram, A., Mackie, K. R., & Stojadinovic, B. (2008). Effect of abutment modeling on the seismic response of bridge structures. *Earthquake engineering and engineering vibration*, *7*(4), 395-402.

Baker, J. W., Luco, N., Abrahamson, N. A., Graves, R. W., Maechling, P. J., & Olsen, K. B. (2014, July). Engineering uses of physics-based ground motion simulations. In *Proceedings of the Tenth US Conference on Earthquake Engineering*.

Boore, D. M. (2003). Simulation of ground motion using the stochastic method. *Pure and applied geophysics*, *160*(3-4), 635-676.

Boore, D. M. (2009). Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bulletin of the Seismological Society of America*, *99*(6), 3202-3216.

Bozorgnia, Y., & Bertero, V. V. (2004). *Earthquake engineering: from engineering seismology to performance-based engineering*. CRC press.

Bradley, B. A., Burks, L. S., & Baker, J. W. (2015). Ground motion selection for simulation-based seismic hazard and structural reliability assessment. *Earthquake Engineering & Structural Dynamics*, *44*(13), 2321-2340.

Bradley, B. A., Razafindrakoto, H. N., & Polak, V. (2017). Ground-motion observations from the 14 November 2016 M w 7.8 Kaikoura, New Zealand, earthquake and insights from broadband simulations. *Seismological Research Letters*, *88*(3), 740-756.

Crempien, J., & Archuleta, R. (2015). UCSB method for broadband ground motion from kinematic simulations of earthquakes. *Seismological Research Letters, 86*(1).

Dreger, D., Beroza, G., Day, S., Goulet, C., Spudich, P., Stewart, J., & Jordan, T. (2015). Evaluation of SCEC broadband platform phase 1 PSA ground motion simulation results. *Seismological Research Lett*ers *86*(1).

Frost, J. (2019). "Standard Error of the Regression vs. R-squared." *Statistics by Jim.*

Galasso, C., Zareian, F., Iervolino, I., & Graves, R. W. (2012). Validation of ground-motion simulations for historical events using SDoF systems. *Bulletin of the Seismological Society of America*, *102*(6), 2727-2740.

Galasso, C., Zhong, P., Zareian, F., Iervolino, I., & Graves, R. W. (2013). Validation of ground-motion simulations for historical events using MDoF systems. *Earthquake Engineering & Structural Dynamics*, *42*(9), 1395-1412.

Ghayoomi, M., & Dashti, S. (2015). Effect of ground motion characteristics on seismic soil-foundation-structure interaction. *Earthquake Spectra*, *31*(3), 1789-1812.

Girden, E. R. (1992). *ANOVA: Repeated measures* (No. 84). Sage.

Graves, R. W. (1998). Three-dimensional finite-difference modeling of the San Andreas fault: source parameterization and ground-motion levels. *Bulletin of the Seismological Society of America*, *88*(4), 881-897.

Graves, R. W., & Wald, D. J. (2004). Observed and simulated ground motions in the San Bernardino basin region for the Hector Mine, California, earthquake. *Bulletin of the Seismological Society of America*, *94*(1), 131-146.

Graves, R., & Pitarka, A. (2014). Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters*, *86*(1), 75-80.

Hartzell, S. H. (1978). Earthquake aftershocks as Green's functions. *Geophysical Research Letters*, *5*(1), 1-4.

Hartzell, S., Harmsen, S., Frankel, A., & Larsen, S. (1999). Calculation of broadband time histories of ground motion: Comparison of methods and validation using strong-ground motion from the 1994 Northridge earthquake. *Bulletin of the Seismological Society of America*, *89*(6), 1484-1504.

Haselton, C. B., Whittaker, A. S., Hortacsu, A., Baker, J. W., Bray, J., & Grant, D. N. (2012, September). Selecting and scaling earthquake ground motions for performing response-history analyses. In *Proceedings of the 15th World Conference on Earthquake Engineering* (pp. 4207-4217). Earthquake Engineering Research Institute.

"How Can I Compare Regression Coefficients Between Two Groups?" UCLA: Institute for Digital Research & Education. https://stats.idre.ucla.edu/spss/faq/how-can-i-compare-regression-coefficients-between-two-groups

Irikura, K. (1983). Semi-empirical estimation of strong ground motions during large earthquakes.

Irikura, K., & Miyake, H. (2011). Recipe for predicting strong ground motion from crustal earthquake scenarios. *Pure and Applied Geophysics*, *168*(1-2), 85-104.

Jennings, P. C., Housner, G. W., & Tsai, N. C. (1968). Simulated earthquake motions.

Kaviani, P., Zareian, F., & Taciroglu, E. (2012). Seismic behavior of reinforced concrete bridges with skew-angled seat-type abutments. *Engineering Structures*, *45*, 137-150.

Lacey, M. 2019. "ANOVA for Regression." *Statistical Topics.* http://www.stat.yale.edu/Courses/1997-98/101/stat101.htm

Lozos, J. C., Olsen, K. B., Brune, J. N., Takedatsu, R., Brune, R. J., & Oglesby, D. D. (2015). Broadband ground motions from dynamic models of rupture on the northern San Jacinto fault, and comparison with precariously balanced rocks. *Bulletin of the Seismological Society of America*, *105*(4), 1947-1960.

McKenna, F., Fenves, G. L., Scott, M. H., & Jeremic, B. (2000). Open system for earthquake engineering simulation (OpenSees). *Univ. of California, Berkeley, CA,< http://opensees. berkeley. edu*.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

Motazedian, D., & Atkinson, G. M. (2005). Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America*, *95*(3), 995-1010.

Olsen, K. B., & Mayhew, J. E. (2010). Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake. *Seismological Research Letters*, *81*(5), 715-723.

Rezaeian, S., & Der Kiureghian, A. (2010). Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthquake Engineering & Structural Dynamics*, *39*(10), 1155-1180.

Rezaeian, S., Zhong, P., Hartzell, S., & Zareian, F. (2015). Validation of simulated earthquake ground motions based on evolution of intensity and frequency content. *Bulletin of the Seismological Society of America*, *105*(6), 3036-3049.

Somerville, P., Collins, N., Abrahamson, N., Graves, R., & Saikia, C. (2001). Ground motion attenuation relations for the central and eastern United States. *US Geological Survey, Award 99HQGR0098, final report*.

Song, S. G., & Somerville, P. (2010). Physics-based earthquake source characterization and modeling with geostatistics. *Bulletin of the Seismological Society of America*, *100*(2), 482-496.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

The main goal of this dissertation is validating simulated ground motions at the engineering demand level. This goal is accomplished by 1) effectively providing a statistical methodology for the characterization of EDP dependence that yields more accurate simulations of EDPs and estimates of loss, and 2) introducing a step by step methodology that can be used to validate simulated ground motions at the structural response level.

In part one of this dissertation, Gaussian copulas are used to characterize the dependence of EDPs which are tested against the current methods suggested by FEMAP-58 Volumes 1 and 2, detailing recommended provisions for performance based seismic assessment of structural systems. Using Gaussian copulas allows for the generation of a suite of EDPs from a smaller number of initial realizations without the need for potentially inaccurate assumptions regarding EDP dependence. Peak floor acceleration and maximum interstory drift ratios for four special steel moment resisting frame buildings are obtained from a large suite of selected and scaled ground motion records. These values of demand are then used to generate EDPs that exhibit the dependence structure of Gaussian copulas and also, that follow a joint lognormal distribution, which follow the current methodology outlined by FEMAP-58 (2015). Results show that EDPs generated using Gaussian copulas are better able to match the population EDPs from which they were sampled, especially with smaller initial input realizations, compared with EDPs generated that follow a joint lognormal distribution.

Furthermore, in terms of the calculation and assessment of loss, results show that for the majority of cases, copula EDPs generate losses that better match population loss compared with

loss generated from joint lognormal EDPs but that joint lognormal loss still matches well with population loss. These results shed light on the availability of more statistical tools that require fewer assumptions that can be implemented to generate EDPs, such as Gaussian copulas. These results highlight the large amount of inherent variability that exists in performance based estimations of economic loss. This dissertation addresses a solution to remove the need for potentially inaccurate assumptions at the EDP level of performance assessment by providing a methodology that is able to minimize the inaccuracies associated with simulating suites of demand sets from a small number of initial analyses. There is a continued need to extend this work to further address room for improvement in the quantification of uncertainty at other levels of performance assessment, for example in terms of damage and loss. This will contribute to a more holistically accurate methodology that will allow for more reliable engineered structures.

In the second part of this dissertation, a methodology, titled Vector Based Intensity-Measure Method (VBIM), is introduced and presented using three case study structures: a 2-story SMRF, 12-story SMRF and a two-span, single column concrete bridge. Using three case study structures, multiple regression analyses are performed to 1) identify the vector of intensity measures that is able to best predict structural response, 2) develop models that predict structural response from both recorded and simulated ground motions, and 3) quantify similarities and differences between simulated and recorded ground motions using the developed models. Results for the 2- and 12-story SMRFs show that the only significant waveform parameter in predicting roof drift is Arias Intensity; for the bridge structure, both Arias intensity and strong motion duration are significant. The regression models (i.e., one based on simulated ground motions and another with recorded ground motions) that estimate structure response to a vector of intensity measures are statistically compared through ANOVA. Results of ANOVA show no significant difference

between regression models of recorded and simulated ground motions out of the four considered simulation methodologies (i.e. Graves and Pitarka, EXSIM, Song, Irikura-Recipe). Through the determination of confidence and prediction intervals of the response based on significant intensity measures, the acceptable range (i.e., tolerance) required in intensity measures in order for simulated ground motions to accurately predict values of response is determined. When considering Arias intensity, there is a much larger interval that encompasses differences in future predictions of building response compared with bridge response for simulated and recorded ground motions. Further, VBIM is applied using ω', an intensity measure that was shown to be insignificant in predicting roof drift for the SMRFs and using *Sa*, an intensity measure that has been proven to be representative of structural response. These applications illustrate the applicability and accuracy of VBIM in identifying the tolerance of response based on different intensity measures.

These results shed light on the similarities between simulated ground motions and their recorded counterparts at the ground motion waveform level and the structural response level. Furthermore, these results and the VBIM methodology can be used to provide ground motion simulation modelers with recommendations regarding the level of accuracy of intensity measures that are needed in order for future estimates of structural response from simulated ground motions to be within an acceptable range of the structural response from recorded ground motions.

APPENDIX A

The results of each multivariate normality test conducted in Chapter 2 are presented here. All outputs are obtained using R statistical analysis software. In Figures A.1, A.4, A.7, A.10, A.13, and A.16, *glp* corresponds to Mardia's multivariate skew statistic, *chi.skew* is the Chi-squared value of the skewness test, *p.value.skew* is the p-value of the skewness statistic, *g2p* is Mardia's multivariate kurtosis statistic, *z.kurtosis* is the z-value of the kurtosis statistic, *p.value.kurt* is the p-value of the kurtosis statistic, *chi.small.skew* is the Chi-squared value of the small sample skewness statistic and *p.value.small* is the p-value of small sample skewness statistic. Both p-values corresponding to kurtosis and skewness should be greater than the significance level (0.05 in this case) in order to confirm multivariate normality of the data. The small p-value and chi-squared are provided for small datasets (n less than 20) which is not the case here as the dataset consists of 100 points.

The outputs of the Henze-Zirkler tests are shown in Figures A.2, A.5, A.8, A.11, A.14 and A.17. The Henze-Zirkler multivariate normality test provides only one test statistic to determine multivariate normality. *HZ* corresponds to value of the Henze-Zirkler test statistic at a significance level of 0.05 and *p-value* is the significance value for the provided HZ statistic.

The outputs of Royston's multivariate normality tests are shown in Figures A.3, A.6, A.9, A.12, A.15, A.18, which similar to the Henze Zirkler test, has one test statistic. *H* corresponds to the value of Royston's H statistic at a significance level of 0.05 and *p-value* is an approximate p-value for Royston's test with respect to equivalent degrees of freedom (edf).

```
   Mardia's Multivariate Normality Test
---------------------------------------
   data : dataset

   g1p            : 62.24796
   chi.skew       : 1027.091
   p.value.skew   : 1.728026e-124

   g2p            : 149.1745
   z.kurtosis     : 17.73936
   p.value.kurt   : 0

   chi.small.skew : 1064.603
   p.value.small  : 2.284567e-131

   Result         : Data is not multivariate normal.

NOTE: For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample s
ize (n) is less than 20 then 'p.value.small' should be used as significance value of skewness instead of 'p.value.skew'.
---------------------------------------
```

Figure A.1: Results output of Mardia's multivariate normality test for 4-story building considering 2% in 50 years hazard

Figure A.2: output of Zirkler's normality story considering years hazard

```
  Henze-Zirkler's Multivariate Normality Test
---------------------------------------------
  data : dataset

  HZ      : 1.269769
  p-va   Royston's Multivariate Normality Test
       ---------------------------------------------
  Resu   data : dataset
------
         H       : 191.0935
         p-value : 5.176387e-40

         Result  : Data is not multivariate normal.
       ---------------------------------------------
```

Results Henze- multivariate test for 4- building 2% in 50

Figure A.3: Results output of Royston's multivariate normality test for 4-story building considering 2% in 50 years hazard

```
   Mardia's Multivariate Normality Test
---------------------------------------
   data : dataset

   g1p            : 50.58618
   chi.skew       : 834.672
   p.value.skew   : 4.982738e-90

   g2p            : 141.9448
   z.kurtosis     : 15.18328
   p.value.kurt   : 0

   chi.small.skew : 865.156
   p.value.small  : 2.207647e-95

   Result          : Data is not multivariate normal.

NOTE: For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample s
ize (n) is less than 20 then 'p.value.small' should be used as significance value of skewness instead of 'p.value.skew'.
---------------------------------------
```

Figure A.4: Results output of Mardia's multivariate normality test for 4-story building considering 10% in 50 years hazard

```
  Henze-Zirkler's Multivariate Normality Test
---------------------------------------------
  data : dataset

  HZ      : 1.122845
  p-value : 9.992007e-16

  Result  : Data is not multivariate normal.
---------------------------------------------
```

Figure A.5: Results output of Henze-Zirkler's multivariate normality test for 4-story building considering 10% in 50 years hazard

```
  Royston's Multivariate Normality Test
-------------------------------------------
  data : dataset

  H       : 173.3257
  p-value : 2.617112e-35

  Result  : Data is not multivariate normal.
-------------------------------------------
```

Figure A.6: Results output of Royston's multivariate normality test for 4-story building considering 10% in 50 years hazard

```
   Mardia's Multivariate Normality Test
---------------------------------------
   data : dataset

   g1p            : 33.16743
   chi.skew       : 547.2627
   p.value.skew   : 1.681131e-42

   g2p            : 125.9818
   z.kurtosis     : 9.539504
   p.value.kurt   : 0

   chi.small.skew : 567.2499
   p.value.small  : 1.408429e-45

   Result          : Data is not multivariate normal.

NOTE: For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample s
ize (n) is less than 20 then 'p.value.small' should be used as significance value of skewness instead of 'p.value.skew'.
---------------------------------------
```

Figure A.7: Results output of Mardia's multivariate normality test for 4-story building considering 50% in 50 years hazard

```
  Henze-Zirkler's Multivariate Normality Test
---------------------------------------------
  data : dataset

  HZ      : 1.237867
  p-value : 0

  Result  : Data is not multivariate normal.
---------------------------------------------
```

Figure A.8: Results output of Henze Zirkler's multivariate normality test for 4-story building considering 10% in 50 years hazard

```
  Royston's Multivariate Normality Test
-------------------------------------------
  data : dataset

  H       : 10.09526
  p-value : 0.05982886

  Result  : Data is multivariate normal.
-------------------------------------------
```

Figure A.9: Results output of Royston's multivariate normality test for 4-story building considering 10% in 50 years hazard

```
   Mardia's Multivariate Normality Test
---------------------------------------
   data : dataset

   g1p            : 131.695
   chi.skew       : 2546.104
   p.value.skew   : 7.163221e-142

   g2p            : 407.8559
   z.kurtosis     : 17.97897
   p.value.kurt   : 0

   chi.small.skew : 2619.414
   p.value.small  : 7.749103e-152

   Result          : Data is not multivariate normal.

NOTE: For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample s
ize (n) is less than 20 then 'p.value.small' should be used as significance value of skewness instead of 'p.value.skew'.
---------------------------------------
```

Figure A.10: Results output of Mardia's multivariate normality test for 8-story building considering 2% in 50 years hazard

```
  Henze-Zirkler's Multivariate Normality Test
---------------------------------------------
  data : dataset

  HZ      : 2.054968
  p-value : 0

  Result  : Data is not multivariate normal.
---------------------------------------------
```

Figure A.11: Results output of Henze Zirkler's multivariate normality test for 8-story building considering 2% in 50 years hazard

```
  Royston's Multivariate Normality Test
-------------------------------------------
  data : dataset

  H       : 55.93793
  p-value : 4.896934e-14

  Result  : Data is not multivariate normal.
-------------------------------------------
```

Figure A.12: Results output of Royston's multivariate normality test for 8-story building considering 2% in 50 years hazard

```
   Mardia's Multivariate Normality Test
---------------------------------------
   data : dataset

   g1p            : 93.31322
   chi.skew       : 1539.668
   p.value.skew   : 9.992152e-29

   g2p            : 359.2756
   z.kurtosis     : 7.100454
   p.value.kurt   : 1.24345e-12

   chi.small.skew : 1591.63
   p.value.small  : 4.60264e-33

   Result          : Data is not multivariate normal.

NOTE: For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample s
ize (n) is less than 20 then 'p.value.small' should be used as significance value of skewness instead of 'p.value.skew'.
---------------------------------------
```

Figure A.13: Results output of Mardia's multivariate normality test for 8-story building considering 10% in 50 years hazard

```
  Henze-Zirkler's Multivariate Normality Test
---------------------------------------------
  data : dataset

  HZ      : 1.021423
  p-value : 0

  Result  : Data is not multivariate normal.
---------------------------------------------
```

Figure A.14: Results output of Henze Zirkler's multivariate normality test for 8-story building considering 10% in 50 years hazard

```
  Royston's Multivariate Normality Test
-------------------------------------------
  data : dataset

  H       : 39.77925
  p-value : 2.991331e-06

  Result  : Data is not multivariate normal.
-------------------------------------------
```

Figure A.15: Results output of Royston's multivariate normality test for 8-story building considering 10% in 50 years hazard

```
   Mardia's Multivariate Normality Test
---------------------------------------
   data : dataset

   g1p            : 116.2324
   chi.skew       : 1917.835
   p.value.skew   : 7.52901e-65

   g2p            : 388.4085
   z.kurtosis     : 12.80282
   p.value.kurt   : 0

   chi.small.skew : 1982.56
   p.value.small  : 5.99153e-72

   Result          : Data is not multivariate normal.

NOTE: For multivariate normality, both p-values of skewness and kurtosis statistics should be greater than 0.05. If sample s
ize (n) is less than 20 then 'p.value.small' should be used as significance value of skewness instead of 'p.value.skew'.
---------------------------------------
```

Figure A.16: Results output of Mardia's multivariate normality test for 8-story building considering 50% in 50 years hazard

```
  Henze-Zirkler's Multivariate Normality Test
---------------------------------------------
  data : dataset

  HZ      : 1.058388
  p-value : 0

  Result  : Data is not multivariate normal.
---------------------------------------------
```

Figure A.17: Results output of Henze Zirkler's multivariate normality test for 8-story building considering 10% in 50 years hazard

```
  Royston's Multivariate Normality Test
-------------------------------------------
  data : dataset

  H       : 10.60255
  p-value : 0.0811815

  Result  : Data is multivariate normal.
-------------------------------------------
```

Figure A.18: Results output of Royston's multivariate normality test for 8-story building considering 50% in 50 years hazard